6-7-2018

Development and Application of Ribose Oxidation Sequencing (RibOxi-seq)

Yinzhou Zhu
University of Connecticut, yinzhou1989@gmail.com

Development and Application of Ribose Oxidation Sequencing (RibOxi-seq)

Yinzhou Zhu, PhD

University of Connecticut, 2018

In eukaryotes, a number of RNA sites are modified by 2'-O methylation (2'-OMe). Such editing is mostly guided by BoxC/D class small nucleolar RNAs (snoRNAs). These snoRNAs direct methylation via complementary RNA-RNA interactions. 2'-OMe has so far been shown to be present in rRNAs, tRNAs and some small RNAs and has been implicated in ribosome maturation and translational circuitries. A substantial portion of known methylated sites in rRNA lie in close proximity to ribosome functional sites such as regions around the peptidyl transferase center. It is not yet clear whether many mRNAs might possess internal 2'-OMe sites. It is therefore important to characterize 2'-OMe landscapes. We have developed a novel method for the highly accurate and transcriptome-wide detection of 2'-OMe sites. The core principle of this method is to randomly digest RNAs to expose 2'-OMe sites at the 3'-ends of digested RNA fragments. Next, an oxidation step using sodium periodate destroys all fragment-3'-ends except those that are 2'-O methylated. Only these oxidation-resistant fragments are available for linker ligation and subsequent sequencing library preparation. We have applied our new method to the study of ribose methylation in both Trypanosoma brucei and Mus musculus liver, and successfully obtained rRNA 2'-OMe profiles for both organisms. Our sequencing data strongly support experimentally validated known sites, while providing several candidates of novel sites, as well as evidence for differential methylation. In conclusion, our method is able to reliably identify 2'-OMe sites in a high-throughput manner. We are also interested in using the method to investigate potential mRNA internal sites, the targets of "orphan" snoRNAs, which have no currently known targets.

Development and Application of Ribose Oxidation Sequencing (RibOxi-seq)

Yinzhou Zhu

B.S., University of Wisconsin-Madison, 2012

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2018

2018

APPROVAL PAGE

Doctor of Philosophy Dissertation

Development and Application of Ribose Oxidation Sequencing (RibOxi-seq)

Presented by

Yinzhou Zhu B.S.

Major Advisor

_____

Gordon G. Carmichael

Associate Advisor

_____

Arthur Gûnzl

Associate Advisor

_____

James Li

Associate Advisor

_____

Sandra K. Weller

Associate Advisor

_____

Stormy Chamberlain

University of Connecticut
2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter I

Introduction

Since establishment of the foundations of molecular biology in the mid-1900s, additional modes of gene regulation, which operate beyond the routes of information flow described in the original central dogma, have been discovered. Layers of complexity to the seemingly simple sequences of bidirectional information transfer from DNA to RNA messages, as well as unidirectional translation from RNA molecules to proteins has been added. While Francis Crick's central dogma of molecular biology remains technically valid, the interconnections between these three biomolecules have transcended generalization that the conventional pathways of information flow. With progress made in the past few decades, we have come to establish a more robust representation of cellular networks, which newly integrates transcriptional and post-transcriptional regulation networks. These two aspects specifically, have grown exponentially partially due to breakthroughs in high-throughput sequencing. Numbers of novel non-coding RNAs such as long non-coding RNAs (lncRNAs) have been identified and functionally characterized, while micro RNAs (miRNAs), small nuclear RNAs (snRNAs), double strand RNA (dsRNA) etc. repertoires have been greatly expanded. More recently, epigenetic modification and epitranscriptomic modification studies have garnered considerable interest and momentum among the scientific field due to their significant implications in transcriptional and post-transcription regulation respectively. Epitranscriptomics, while significantly younger than most other fields, encompasses RNA modifications that have been known for a very long time. The evolution of collective knowledge of each known RNA modification, which I will discuss in greater details later in the current chapter, suggests

increasing scientific realization that these modifications, albeit 'old', might still possess important and novel insights into biological processes and regulation.

A. Emergence of epitranscriptomics.

Epigenetics. In 1906, when William Bateson first referenced very little-known *Principals of Heredity* written by Gregor Mendel, the era of modern genetics began with study of monogenic traits, or in other words: Mendelian traits. The inheritance unit of the modern genetics is defined by 'gene', which was introduced by Danish botanist Wilhelm Johannsen, where disparities between copies of a single gene are inheritable and contributes to trait differential displayed on different individuals within the same organism (Johannsen 1905). Later, contrary to the classic mendelian inheritance, the patterns of inheritance for many traits were discovered to follow no particular criteria; some traits seemingly mapped to only single genes that violate Mendelian rules, or traits having unexplained inheritance plasticity, giving rise to the idea of non-Mendelian inheritance. Studies have then established as complex traits, where multiple genes, sex-linked traits or incomplete dominance/codominance come into play. (reviewed in Correns 1937, Kowles 2001, van Heyningen 2004, Cuzin 2013). With advances in biological sciences and technology, genetics now deals with finer resolution. We are now looking at genes under single nucleotide level and studying how change in bases affect gene expression levels, thus affecting the phenotypes of cell types, organs and ultimately organisms. Although the early models of inheritance still apply, they fail to account for inheritable traits or gene expression level changes not associated with changes in actual DNA sequences. It was not until 1942, when Waddington showed inheritable characteristics of *Drosophila* reacting to environmental stimuli, the study of epigenetics ("above" genetics) was proposed (Waddington 1942). Subsequently, mechanisms such as DNA methylation, in the form of 5-methylcytosine (m^5C), was shown to regulate vital

cellular processes including X inactivation, and differentiation related gene expression levels (reviewed in Felsenfeld 2007). Both DNA and chromatin methylation and demethylation since then been implicated in the mechanism of epigenetics and incorporated into the field of study, and epigenetics deals with a collection of modifiers that modulate gene expressions without changing nucleotide base identity. These epigenetic modifications are categorized into several types including DNA (mainly in the form of methylation) and histone (methylation, acetylation and phosphorylation) modifications, each with their own sets of modifying enzymes (reviewed in Gardner et al. 2011, Bird 2002). The growth of the epigenetics field, especially for DNA modifications, was accelerated with the advent of the next generation sequencing technologies (NGS), which made high-throughput identification available through combining NGS and bisulfite treatment. Such breakthroughs allowed us to further the study of the field, which leading to our current expansive insights into epigenetic regulatory networks that are involved in not only basic concept of inheritance, but also in pathology and oncology of various disorders (reviewed in Basse & Arock 2015, Mummaneni & Shord 2014, Ordog et al. 2012).

Epitranscriptomics. Similar to DNA, RNA transcripts are replete with chemical modifications. Unlike epigenetics, which revolves around just a few types of marks, the RNA modification repertoire is highly diverse in terms of chemical properties (reviewed in He 2010). Up until the early 2010s, most important known RNA modifications to date were on tRNAs, small RNAs and rRNAs, although most of these modifications have been chemically characterized for decades, how they function was largely unclear (Motorin & Grosjean 1998). While the most recent curation has revealed approximately 170 different types of RNA modifications, novel modifications are still being actively discovered and added. Also, the estimated number of sites across the transcriptome could be far more than just a few hundreds

(Boccaletto et al. 2018). Contrary to epigenetic marks, whose regulatory roles on gene expressions are carried out by altering accessibility of genomic regions, RNA modifications have been traditionally regarded as structural features that affect transcript rigidity and stability (Sloan et al. 2017). While the collective significance of them is still being actively pursued, regulatory functions have been ascribed for some RNA editing, including Apobec3 induced C to U deamination, ADAR induced A-to-I editing, and N6-methyladenosine (m^6A). The collective evidence and evolution of methods to study RNA modifications have thus resulted in the emergence of the field of transcriptomics, where chemical changes/additions to the RNA that confer additional layers of biological functions are examined. Most of the modifications already known for decades might not seem appealing when looked at through conventional methods, but when coupled with recent advancements in NGS technologies, these modifications are garnering renewed interest and optimism through the field of epitranscriptomics as a result of continuing discoveries of their novel roles.

RNA base methylations. Modifications on RNA transcripts form a highly distinct set. The modifications can happen in a variety of ways. For example, a methyl group can be added to an adenosine at different positions on the nitrogenous base to form 1-methyladenosine (m^1A), 2-methyladenosine (m^2A) etc (Dunn 1961, Starr & Fefferman 1964). A different base such as cytosine, can also be modified with the same chemical group; In addition to structurally distinctive modifications with the same chemical group, more than 50 different chemical moieties are contained in the set including thiols, esters etc; all modifications but a few, which occur on the ribose of a nucleotide (2'-O-methyl, 2'-O-ADP etc.), are found on the nitrogenous base of the nucleotides (base methylation, reviewed in Wang & He 2014). Very interestingly, RNA such as mRNA and tRNA can contain more than 50 different base modifications, while 17

occur on eukaryotic rRNA alone. Because many of them have been extensively characterized across different species spanning all 3 domains of life, phylogenetic studies have been possible, revealing that RNA modifications are conserved (Cantara et al. 2011, and references therein). A table of prevalent modifications summarized from Cantara et al. (2011) and Pietro et al. (2017) is provided here (Table 1). Although the number of modifications is large, there is a core subset of them that have been focused on by researchers due to functional importance, prevalence of the modification and overall distribution patterns. To have an accurate reflection of the recent status of the field of epitranscriptomic studies, I will focus on the following RNA modifications: N^6-methyladenosine (m^6A), 5-methylcytosine (RNA: m^5C), N^1-methyladenosine (m^1A), Adenosine to Inosine editing (A-to-I), pseudouridine (Ψ), and 2'-O-methylation (2'-OMe).

Table 1. List of major RNA modifications for Archaea, Bacteria and Eukarya.

Symbol	Common Name	Found on:
m^3Um	3,2′-O-dimethyluridine	mRNA
$m^{2,2,7}G$	N2, N2,7-trimethylguanosine	mRNA,snRNA
m^6Am	N6,2′-O-dimethyladenosine	mRNA,snRNA
$m^{2,2}G$	N2, N2-dimethylguanosine	rRNA
m^4C	N4-methylcytidine	rRNA
m^1acp^3Y	1-methyl-3-(3-amino-3-carboxypropyl) pseudouridine	rRNA
m^3Y	3-methylpseudouridine	rRNA
m^3U	3-methyluridine	rRNA
cm^5U	5-carboxymethyluridine	rRNA
hm^5C	5-hydroxymethylcytidine	rRNA
m^8A	8-methyladenosine	rRNA
Nm	2′-O-methyladenosine	rRNA,tRNA,snRNA
	2′-O-methylcytidine	
	2′-O-methylguanosine	
	2′-O-methylinosine	
	2′-O-methylpseudouridine	
	2′-O-methyluridine	
$m^{6,6}Am$	N6, N6,2′-O-trimethyladenosine	rRNA,mRNA
m^2G	N2-methylguanosine	rRNA,snRNA
m^1A	1-methyladenosine	rRNA,tRNA
m^1G	1-methylguanosine	rRNA,tRNA
m^1Y	1-methylpseudouridine	rRNA,tRNA
m^2A	2-methyladenosine	rRNA,tRNA
acp^3U	3-(3-amino-3-carboxypropyl) uridine	rRNA,tRNA
m^3C	3-methylcytidine	rRNA,tRNA
m^5Um	5,2′-O-dimethyluridine	rRNA,tRNA
$m^{4,4}Cm$	N4, N4,2′-O-trimethylcytidine	rRNA,tRNA

ac^4Cm	N4-acetyl-2′-O-methylcytidine	rRNA,tRNA
ac^4C	N4-acetylcytidine	rRNA,tRNA
m^7G	7-methylguanosine	rRNA,tRNA,mRNA
m^6A	N6-methyladenosine	rRNA,tRNA,mRNA,snRNA
□	pseudouridine	rRNA,tRNA,mRNA,snRNA
D	Dihydrouridine	rRNA,tRNA,snRNA
m^1Nm	1,2′-O-dimethyladenosine	tRNA
	1,2′-O-dimethylguanosine	
	1,2′-O-dimethylinosine	
m^1I	1-methylinosine	tRNA
$msms^2i^6A$	2- methylthiomethylenethio-N6-isopentenyl-adenosine	tRNA
k^2C	2-lysidine	tRNA
ms^2io^6A	2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine	tRNA
ms^2hn^6A	2-methylthio-N6-hydroxynorvalylcarbamoyladenosine	tRNA
ms^2m^6A	2-methylthio-N6-methyladenosine	tRNA
ms^2t^6A	2-methylthio-N6-threonylcarbamoyladenosine	tRNA
s^2Um	2-thio-2′-O-methyluridine	tRNA
s^2C	2-thiocytidine	tRNA
s^2U	2-thiouridine	tRNA
Ar(p)	2′-O-ribosyladenosine (phosphate)	tRNA
Gr(p)	2′-O-ribosylguanosine (phosphate)	tRNA
$imG\text{-}1^4$	4-demethylwyosine	tRNA
s^4U	4-thiouridine	tRNA
m^5Cm	5,2′-O-dimethylcytidine	tRNA
$mchm^5Um$	5-(carboxyhydroxymethyl) -2′-O-methyluridine methyl ester	tRNA
inm^5s^2U	5-(isopentenylaminomethyl) -2-thiouridine	tRNA

inm^5Um	5-(isopentenylaminomethyl)-2′-O-methyluridine	tRNA
inm^5U	5-(isopentenylaminomethyl)uridine	tRNA
nm^5se^2U	5-aminomethyl-2-selenouridine	tRNA
nm^5s^2U	5-aminomethyl-2-thiouridine	tRNA
ncm^5Um	5-carbamoylmethyl-2′-O-methyluridine	tRNA
ncm^5U	5-carbamoylmethyluridine	tRNA
$cmnm^5s^2U$	5-carboxymethylaminomethyl-2-thiouridine	tRNA
$cmnm^5Um$	5-carboxymethylaminomethyl-2′-O-methyluridine	tRNA
$cmnm^5U$	5-carboxymethylaminomethyluridine	tRNA
f^5Cm	5-formyl-2′-O-methylcytidine	tRNA
f^5C	5-formylcytidine	tRNA
ho^5U	5-hydroxyuridine	tRNA
mcm^5s^2U	5-methoxycarbonylmethyl-2-thiouridine	tRNA
mcm^5Um	5-methoxycarbonylmethyl-2′-O-methyluridine	tRNA
mcm^5U	5-methoxycarbonylmethyluridine	tRNA
mo^5U	5-methoxyuridine	tRNA
m^5s^2U	5-methyl-2-thiouridine	tRNA
mnm^5se^2U	5-methylaminomethyl-2-selenouridine	tRNA
mnm^5s^2U	5-methylaminomethyl-2-thiouridine	tRNA
mnm^5U	5-methylaminomethyluridine	tRNA
tm^5s^2U	5-taurinomethyl-2-thiouridine	tRNA
tm^5U	5-taurinomethyluridine	tRNA
m^2Gm	N2,2′-O-dimethylguanosine	tRNA
$m^{2,7}Gm$	N2,7,2′-O-trimethylguanosine	tRNA
$m^{2,2}Gm$	N2, N2,2′-O-trimethylguanosine	tRNA
io^6A	N6-(cis-hydroxyisopentenyl) adenosine	tRNA
ac^6A	N6-acetyladenosine	tRNA
g^6A	N6-glycinylcarbamoyladenosine	tRNA

hn^6A	N6-hydroxynorvalylcarbamoyladenosine	tRNA
i^6A	N6-isopentenyladenosine	tRNA
m^6t^6A	N6-methyl-N6-threonylcarbamoyladenosine	tRNA
t^6A	N6-threonylcarbamoyladenosine	tRNA
OHyW	hydroxywybutosine	tRNA
OHyWy	methylated undermodified hydroxywybutosine	tRNA
mimG	methylwyosine	tRNA
o^2yW	peroxywybutosine	tRNA
cmo^5U	uridine 5-oxyacetic acid	tRNA
$mcmo^5U$	uridine 5-oxyacetic acid methyl ester	tRNA
imG	wyosine	tRNA
Q	queuosine	tRNA
oQ	epoxyqueuosine	tRNA
galQ	galactosyl-queuosine	tRNA
manQ	mannosyl-queuosine	tRNA
preQ0	7-cyano-7-deazaguanosine	tRNA
preQ1	7-aminomethyl-7-deazaguanosine	tRNA
I	inosine	tRNA, mRNA
m^5C	5-methylcytidine	tRNA, mRNA
$m^{2,7}G$	N2,7-dimethylguanosine	tRNA, mRNA

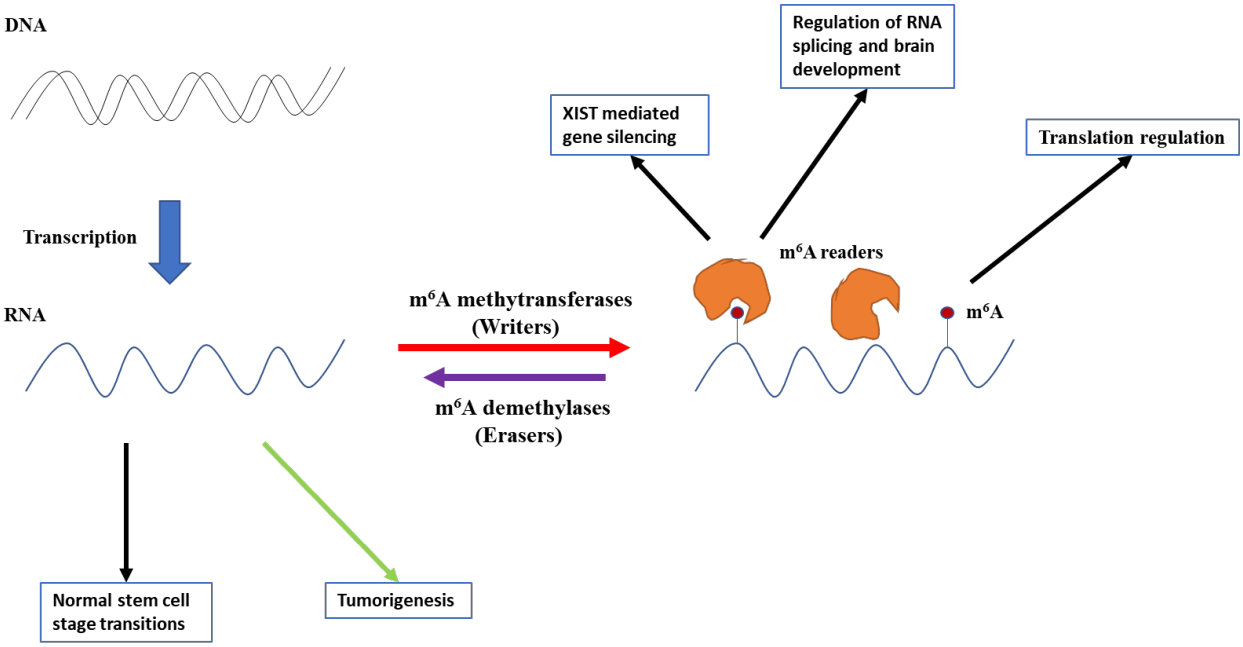
M^6A is the most abundant base methylation epitranscriptomic mark known to date. This modification was originally discovered in tRNAs rRNAs and poly-A RNA fractions (Bokar et al. 1997). M^6A was originally thought to modulate tRNA and rRNA folding structures and stabilities due to its base pairing weakening properties. Although *in vitro* biochemical techniques in earlier studies that used synthetic oligos, as well as C^{14} methionine labeling techniques, gave insights of the potential functions of the modifications, the physiological relevance of these *in vitro* results was not established (reviewed in Saletore et al. 2012). At the same time, mapping sites had proven to be difficult and laborious using low-throughput HPLC or thin-layer chromatography; most frustratingly, it was hard to determine where to look in longer transcripts such as rRNAs and mRNAs (reviewed in Zhao et al. 2016). Thus, progress on the 5-decade-old modification remained stagnant. However, emergence of NGS technology brought capability of high-throughput DNA sequencing, which was then developed to also support RNA sequencing. It did not take long before the next gen RNA sequencing technology was adapted for high-throughput m^6A detection. Dominissini et al. (2012) and Meyer et al. (2012) took advantage of the existence of m^6A antibodies to immunoprecipitate RNAs with the modification followed by subjecting the RNA fragments to RNA sequencing procedures after fragmentation. Subsequent alignment of the sequenced fragments would then allow assigning intervals where RNAs are modified down to 25-100 base pair resolution transcriptome-wide. Due to the nature of the immunoprecipitation, it was not possible to obtain single-base resolution. However, this experimental design allowed researchers to finally generate m^6A profiles, to perform association studies, and to eventually narrow in on genes/pathways of interest followed by validation using traditional methods. Since then a number of biological pathways involving the modification have characterized, for example the findings that: naturally occurring modifications with m^6A affect

levels of pluripotency-promoting transcripts through transcriptional regulation. Thus disrupting steady-state demethylation alters embryonic stem cell (ESC) transition from the naïve to the primed state, while hyper demethylation contributes to tumorigenesis (Cui et al. 2017); m^6A , highly enriched in the brain, can also regulate splicing, and is involved in brain development, sex determination in *Drosophila* via alternative splicing, as well as human x chromosome inactivation(Lence et al. 2016, Patil et al. 2016, Xiao et al. 2016); and the modification has also been recently shown to directly impact mRNA translation elongation through altered base pairing thermal dynamics with tRNAs (Choi et al. 2016). Owing to the now expanded lists of m^6A writers (methyltransferases), erasers (demethylases), readers (m^6A binding proteins) and clear biological functions characterized (RNA stability, signaling and splicing), m^6A is the epitome of the field of epitranscriptomics with constructions of m^6A landscapes playing vital roles in discoveries/validations of specific concepts and results (Figure 1, reviewed in Niu et al. 2012, Zhao et al. 2016).

Figure 1. Functional Roles of m^6A

A simple diagram summarizing m^6A mediated functions. Black arrows represent functions mediated under nominal conditions. The green arrow represents methylation/demethylation imbalance, resulting in decrease in overall methylation level.

Diverse functions of RNA modification by means of N6-methyladensine



Both m^5C and m^1A are less well characterized epitranscriptomic marks compared to m^6A. Similar to m^6A, m^5C and m^1A are RNA modifications described many years ago (Yuki & Fujiwara 1976, Hong et al. 1997). Like its sibling DNA m^5C, RNA m^5C had been shown to be potentially present on polyadenylated RNA decades ago, when it was discovered; although possessing almost the same chemical properties as m^5C in DNA, the biological functions of RNA m^5C's remain largely elusive (reviewed in Oerum et al. 2017). Unlike m^6A, which has been extensively mapped across different RNAs, cell types and species, m^5C has only a few available transcriptome wide mapping results available for further study to date. Bisulfite sequencing, which is the high-throughput detection method available for both DNA and RNA, has been available and became a rather mature detection method (Schaefer et al. 2009). Thus, the pace of m^5C profiling has been picking up recently, and studies have started filling in some gaps including: 1) analysis of m^5C conservation in plant rRNAs (Burgess et al. 2015); 2) differential methylation studies between mouse organs (Johnson et al. 2016); 3) evidence of m^5C shaping chromatin organization, affecting drug resistance to leukemia(Cheng et al. 2018); and 4) correlation of 5hmC, a derivative of m^5C, with global modification patterns and glioblastoma survival (Amort et al. 2017 and references therein). M^1A on the other hand, was identified as a novel modification on poly(A)$^+$ RNAs. It was only detected in mRNAs and some long non-coding RNAs in 2016 after antibody-based IP coupled with NGS was adapted from m^6A detection method (Dominissini et al. 2016). Realization of the importance of high-throughput detection techniques has led to the active development of more methods for its detection and mapping (Ritchey et al. 2017). Although, available datasets for both m^5C and m^1A mainly point to structural roles contributing to RNA stability, further probing and validations are required to examine whether these two modifications could have similar levels of functions to m^6A. We

expect to gain more insights as more data are generated from current methods and methods being actively developed.

Guide RNA dependent modifications. m^6A, RNA m^5C and m^1A are added to target RNA bases through guidance to specific RNA motifs by other protein components that directly recognize the motifs. On the other hand, A-to-I editing and pseudouridylation are guide RNA dependent. A-to-I RNA editing is the chemical reaction where adenosine undergone hydrolytic deamination by adenosine deaminases that act on RNA (ADAR) protein family members, resulting in inosine, which base-pairs with C instead of U and functions genetically and biochemically in the cell as guanosine (Polson et al. 1991). This is one of the special cases where modification changes the nucleotide identity; however, as the field of epitranscriptomics evolves, the definition has been expanded to include modifications confer regulatory functions. And the current consensus is that the A-to-I editing can be considered as a type of epitranscriptomic change (reviewed in O'Connell 2015). A-to-I editing, which has both nuclear and cytoplasmic components, occurs on dsRNAs including sense-anti-sense paired transcripts and intramolecularly paired repeat elements such as *Alu*, SINE and LINE elements; editing sites of ADARs were originally discovered through sequence mismatches when comparing genomic sequences with cDNA sequences from particular RNA transcripts. ADAR editing has been reported to regulate miRNA processing and efficacy in cytoplasm; it has also been shown that elicitation of interferon responses against dsRNAs are suppressed by ADAR editing (reviewed in Nishikura 2010). It was not until NGS became available that A-to-I editing sites could be analyzed on a large scale (reviewed in Nishikura 2016). From past accumulated evidence together with recent deep sequencing studies, we now have a more comprehensive understanding of the actions of ADAR family proteins. It has been demonstrated from the collective profiles of

the editing sites that ADAR proteins exhibit two modes of activity – promiscuous editing and site selective editing. As a result of promiscuous editing, a significant portion of the human transcriptome has extensive stretches of A to I conversions, with *Alu* RNAs being the majority. It was shown to have vital role in preventing cellular damage from dsRNA mediated innate responses. Such function is highly correlated with inhibition of nuclear export of the highly edited RNA transcripts into the cytoplasm (reviewed in Nishikura 2010, O'Connell et al. 2015). Site-selective mode, on the other hand, pinpoint and modify a small group of adenosines resulted from limited imperfect base-pairing of neighboring bases. Such editing mode has been reported to occur on several mRNA transcripts, including AMPA receptor subunit GluR-B, K(V) potassium channel etc. The editing on the GluR-B, a glutamate gated ion channel, was shown to be important for brain functions (Higuchi et al. 1993, Greger et al. 2002, Greger et al. 2003). It was also reported that editing on K(V) mRNA is a mechanism of controlling channel inactivation (Bhalla et al. 2004).

ADAR editing is also abundantly enriched during viral infections, such as by Murine polyomavirus (MPyV). During infection, overlapping portions of viral RNA concatemers are promiscuously edited (Chen & Carmichael 2009, Garren et al. 2015). Consistent with an *in vitro* study, where viral gene expressions were downregulated by editing, the RNA modification can have additional role in anti-viral responses (Gu et al.2009). Although such observations indicate certain roles the modification play during said biological processes, whether they are the cause or result are still being debated and will need to be verified. It is encouraging to see that creatives tool have been actively developed for the study of ADAR editing: CRISPR guide RNAs are introduced to form double stranded structures with targets, while ADAR is tethered to the dCas9 for targeted editing (Fukuda et al. 2017). Such strategy would allow generation of related tissue

cultures and transgenic mice lines for easier study of the editing under more physiological relevant conditions. While we wait for more studies to come out, clinical applications of the ADARs are also being explored.

Recent methodological advancements have also enabled thorough profiling of RNA pseudouridylation, which was the first RNA modification discovered and which results in the conversion of uridine to pseudouridine through RNA guided isomerization of the uridine by pseudouridine synthase (PUS) (Massenet et al. 1999). Following early discoveries made on tRNAs and rRNAs, the high density of this modification was generally believed to be the most abundant cellular RNA modification (Cohn and Volkin 1951). The RNA guiding the modification is a class of small nucleolar RNA (snoRNA) that possesses RNA motifs called box H/ACA motifs. snoRNAs are assembled into ribonucleoprotein particles (snoRNPs) that possess both targeting and enzymatic functions. The H and ACA motifs together with neighboring sequences determine target RNA sites, while scaffolding proteins and PUS are assembled around the boxes (reviewed in Li et al. 2016). Pseudouridylation patterns have been shown to be highly conserved in rRNAs, tRNAs and small nuclear RNAs (snRNAs). Pseudouridines have been considered to mainly contribute to rRNA folding and occur in a conserved pattern around the peptidyl transferase center (ptc) -interacting region (reviewed in Ge & Yu 2013). Pseudouridine, together with 2'-O methylation both utilize the snoRNA machinery. Although two modifications are chemically different, they conventionally seem to have similar functions. Importantly, a transcriptome-wide deep sequencing detection method (Pseudo-seq) using a chemical modifier specific to pseudouridine was recently developed to identify pseudouridines at single-base resolution (Schwartz et al. 2014, Carlile et al. 2014). Since then, additional novel pseudouridylation sites in both non-coding RNAs and mRNAs have been identified, at the same

time, this could be where pseudouridylation and 2'-O methylation diverge. Besides expanding the pseudouridylation landscape, pseudouridine profiles generated so far has shown some potential correlations between the modification and diseases. For example, normal and dyskeratosis congenita patient pseudouridylation patterns were profiled and results suggested strong enrichment and differential modification on telomerase RNA component (TERC), which is implicated in the pathology (reviewed in Zhao & He 2015). Li et al. (2016) have noted a renewed interest and speculated the potential role of pseudouridine in gene regulation.

Although there are currently a number of other examples of RNA modification, such as C-to-U editing by APOBEC family protein members (which have also been identified as DNA modifiers), I feel the critical point to emphasize is the importance of development of new methodologies that allow the of biological processes/components and discoveries of new applications has been properly expressed. In the rest of this thesis, I will focus on a single RNA modification, 2'-O-methylation.

B. Ribose methylation/2'-O-methylation.

Mechanisms of snoRNA guided 2'-O-methylation. 2'-O-methylation is the addition of a methyl group to the 2'-O-position of a nucleotide sugar backbone-ribose-, which is why it is also referred to as ribose methylation (Figure 2A). Deposition of the modification is an RNA-guided event similar to that of the RNA pseudouridylation, where snoRNAs play a central role in the process. Contrary to pseudouridylation, ribose methylation is guided by snoRNAs that contain box C/Ds instead of box H/ACAs, despite the existence of rare snoRNAs that can guide both 2'-O-methylation and pseudouridylation.

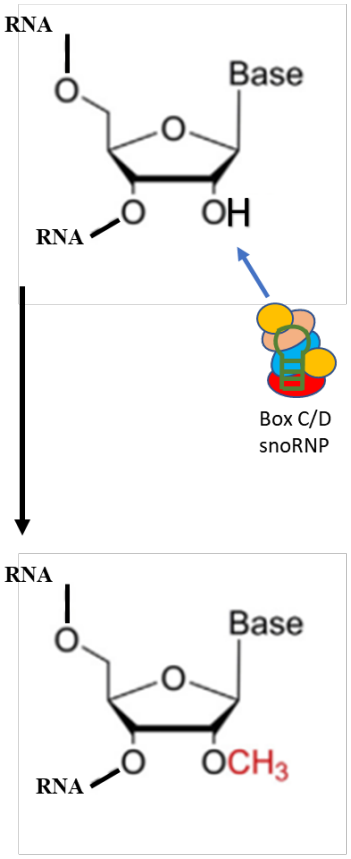
Figure 2. Box C/D snoRNA guided 2'-O-methylation and Box C/D snoRNP biogenesis

(A) A simplified diagram for box C/D snoRNA guided 2'-O-methylation. The addition of one methyl group, which is highlighted in red on the bottom panel, to the ribose 2'-O position is carried out by box C/D snoRNPs.

(B) Box C/D snoRNP biogenesis for human is summarized here. snoRNA biogenesis for other eukaryotes and Archaea undergo same procedures, while Nop56, Nop58, 15.5k and fibrillarin are replaced with other protein homologs. Box H/ACA snoRNPs biogenesis also share the same mechanism with box C/D snoRNPs with the exception of recruiting different scaffolding and catalytic proteins.

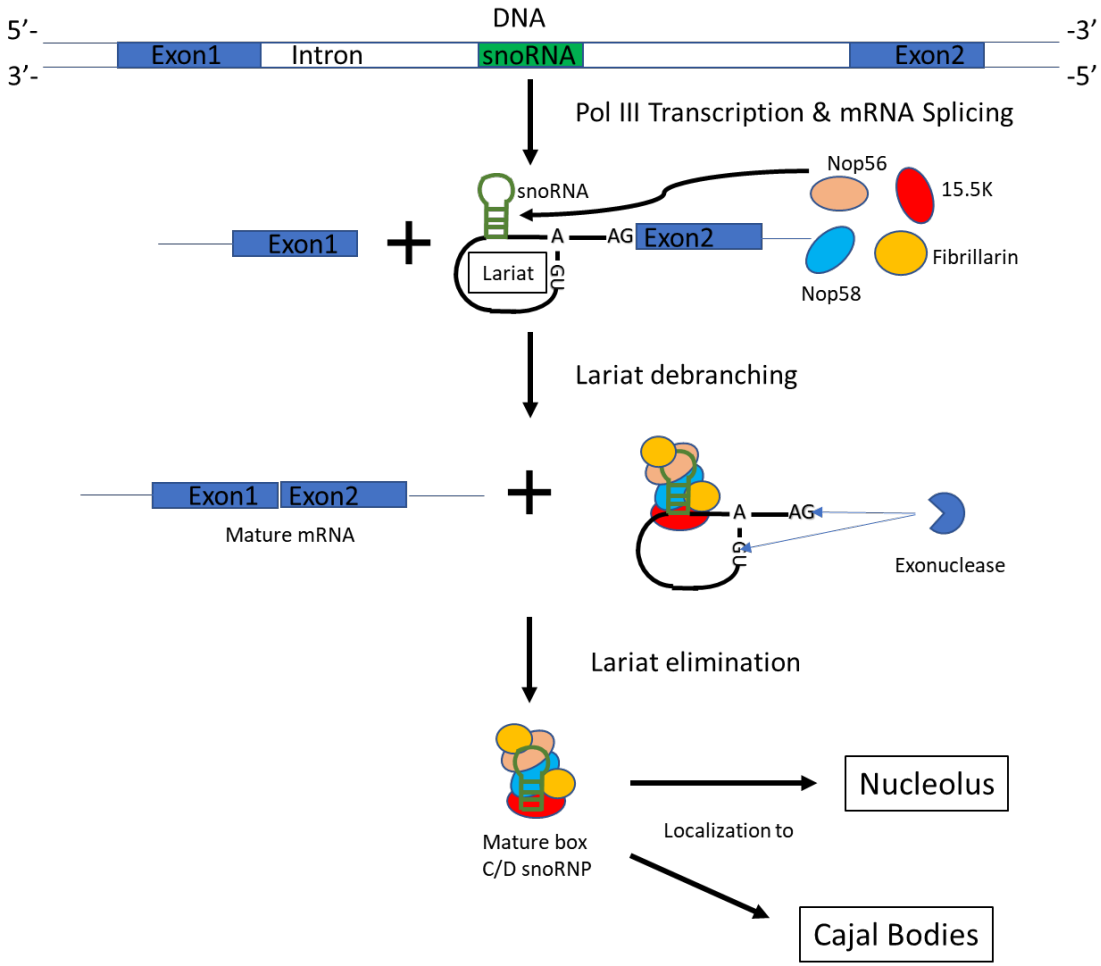
Box C/D snoRNA mediated 2'-O-methylation

A



Box C/D snoRNP biogenesis

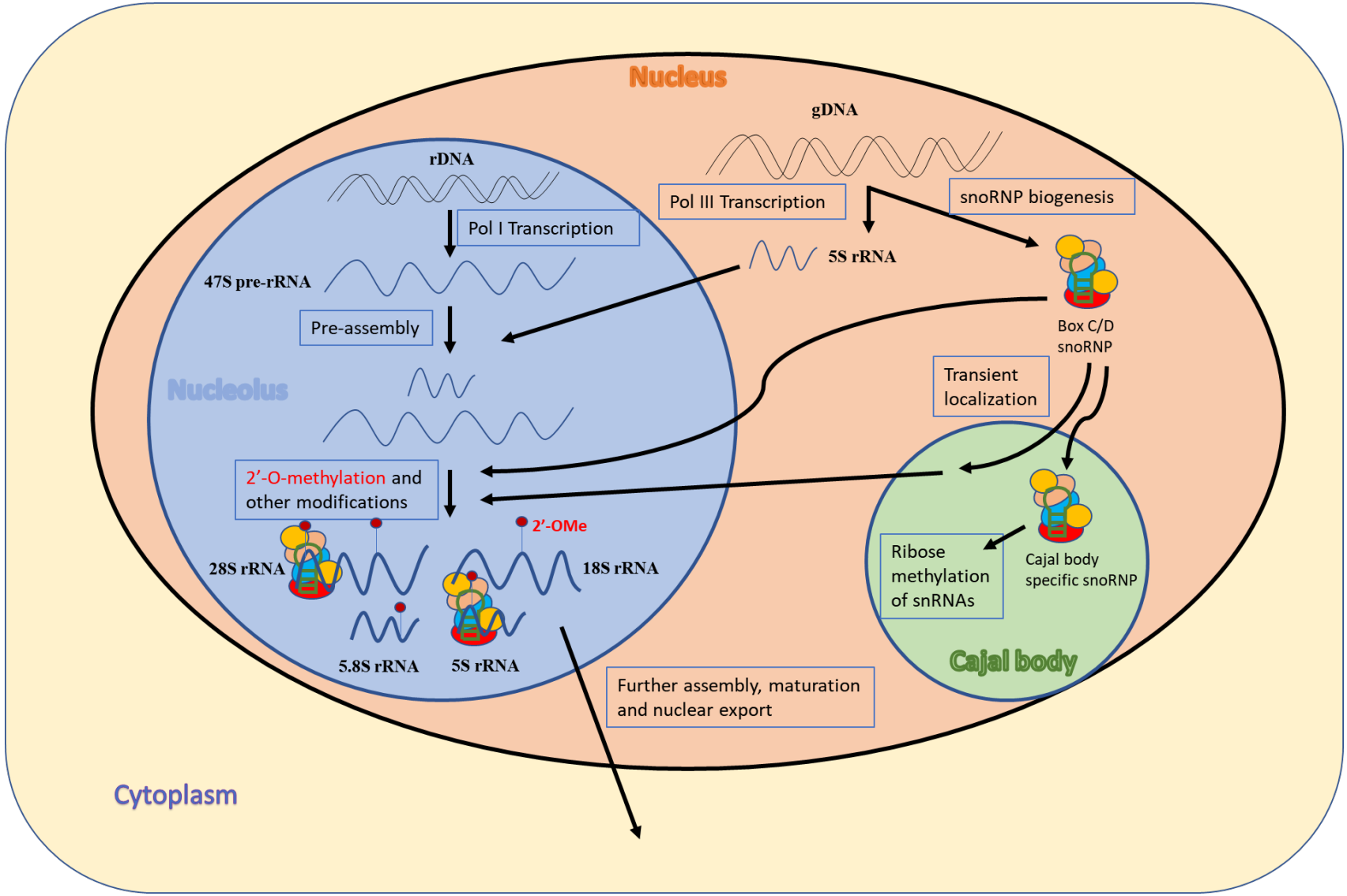
B



Box C/D snoRNAs (H/ACA biogenesis is similar) are processed inside the nucleus from intronic regions of RNA PolII transcribed host RNAs that undergo splicing. Between steps of spliceosome assembly to lariat formation during co-transcriptional splicing of the host RNAs, scaffolding proteins 15.5K, Nop58, Nop56 and the methyltransferase Fibillarin become associated with the C/D boxes (reviewed in Kiss 2006). When the lariat is debranched, exonuclease degradation occurs to eliminate the entire intron except for the snoRNA portion protected by the bound proteins, forming functional snoRNPs (Figure 2B, Baserga et al. 1991, Hirose et al. 2003). Some snoRNPs permanently localize to Cajal bodies and are subsequently named small Cajal body-specific RNPs (scaRNPs). These RNPs carry out modifications of small nuclear RNAs (snRNAs). Some others transiently localize to Cajal bodies before eventual trafficking into the nucleolus, while the rest localize directly to the nucleolus (Kiss et al. 2006). These RNPs directly participate in the maturation of ribosomal RNAs by means of extensive methylation prior to the assembly of ribosomal subunits and processed rRNAs inside the nucleolus (Figure 3, reviewed in Pelletier et al. 2017). It has been known that tRNAs are also internally methylated, while miRNAs are 2'-O-methylated at their 3' ends. These RNAs are methylated through site-specific protein-mediated processes (reviewed in Clouet-D'Orval et al. 2005). Interestingly, while there is no evidence indicating eukaryotic tRNA and miRNA methylation through snoRNPs, both protein-mediated and guide RNA-dependent mechanisms occur on Archaea tRNAs (Renalier et al. 2005).

Figure 3. Box C/D snoRNP trafficking and its roles in RNA processing

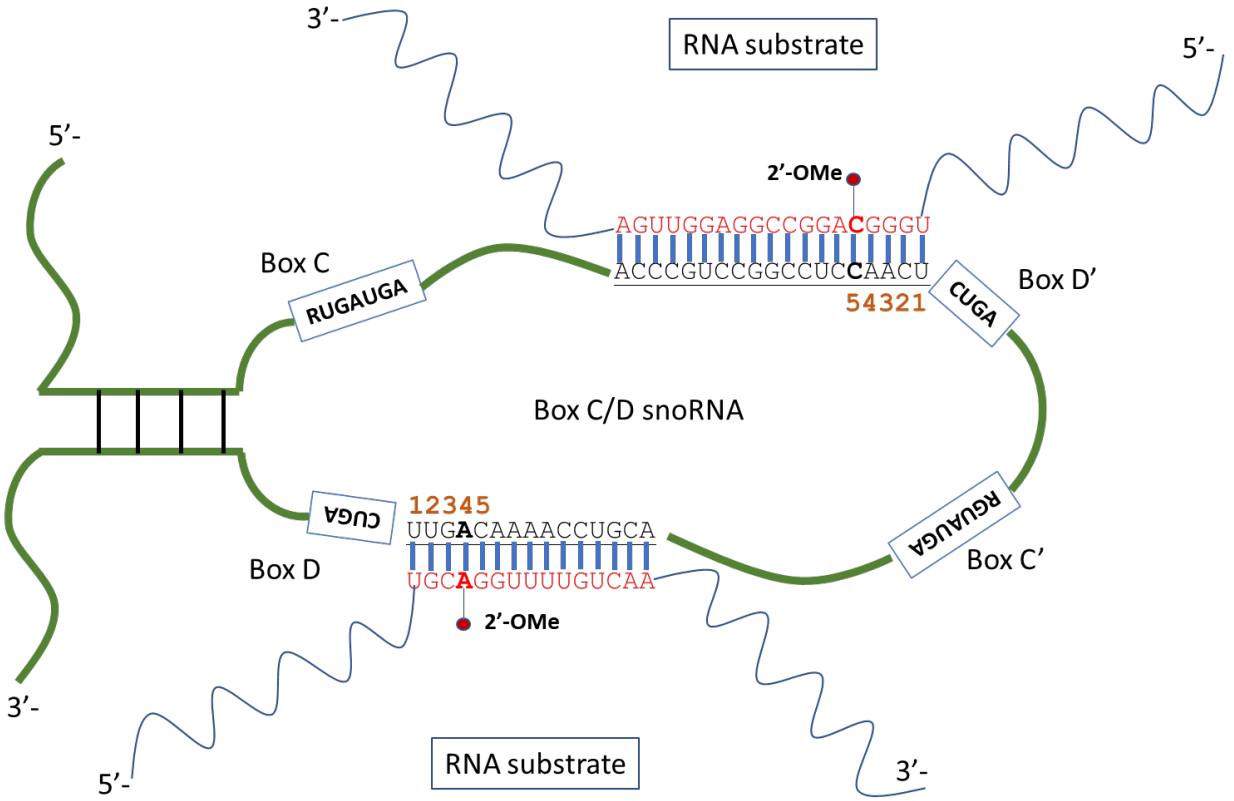
Box C/D snoRNPs either directly localize to nucleolus, or transiently travel to cajal bodies before localization to nucleolus. Box C/D snoRNPs carry out 2'-O-methylation on pre-rRNAs, which is a necessary step for rRNA maturation and ribosome biogenesis. The snoRNPs also 2;-O-methylate snRNAs inside Cajal bodies.


Nucleus
gDNA
rDNA
Pol III Transcription
snoRNP biogenesis
Pol I Transcription
5S rRNA
47S pre-rRNA
Box C/D
snoRNP
Pre-assembly
Transient
localization
Nucleolus
2'-O-methylation and
other modifications
2'-OMe
28S rRNA
18S rRNA
Ribose
methylation
of snRNAs
Cajal body
specific snoRNP
5.8S rRNA
5S rRNA
Cajal body
Further assembly, maturation
and nuclear export
Cytoplasm

Box C/D snoRNAs fold into stem loop structures but retain spacers of at least 12-bp on each side of the stem loop (between Box C and Box D', and between Box C' and Box D), and these anneal to target substrates; the spacer sequences form a maximum of 10-bp complementary base pairs with targets in Archaea, while lengths of duplexes are rather variable in eukaryotes (Kiss 2001, Yang et al. 2016). In most cases, methylation, catalyzed by fibrillarin, occurs at the 4th or 5th base in the base pairing region upstream of box D and/or D', and mismatches around the region abolish methylation activity (Figure 4, Cavaillé, & Bachellerie 1998). 2'-OMes densely populate rRNAs and possibly other RNAs, necessitating a number of different snoRNAs guiding the extensive modifications. Interestingly, cases where one snoRNA can mediate the methylation of multiple sites in human, such as U24 snoRNA on 28S rRNA base location 2350 and 2364, have been observed possibly due the repetitive nature of some regions of rRNAs. Not surprisingly, box H/ACA snoRNAs can also display such a property including ACA17 snoRNA, which guides pseudouridylation on 28S rRNA at both base position 4678 and 4956. In some rare cases, different snoRNAs can target the same site. Taking U51 and U32A for example, each can target base position 1523 on 28S rRNA, but at the same time, each has a different target at a different position (https://www-snorna.biotoul.fr/browse.php?sno=CDBox). Whether this redundancy is selected for during evolution or just a coincidental remains unclear. It is possible that 2'-OMe on 28S rRNA position 1523 contributes more either structurally or functionally, studies mutating the site or probing structural properties might provide some answers.

Figure 4. Typical box C/D snoRNA guide and target

Box C/D snoRNA sequences immediately upstream of box D or box D' base pair with target RNA to form duplex structures. In most cases, either the 4th or the 5th base upstream of the two boxes is methylated.


3'-
RNA substrate
5'-
5'-
2'-OMe
AGUUGGAGGCCGGACGGGU
ACCCGUCCGGCCUCCAACU
54321
Box C
RUGAUGA
CUGA
Box D'
Box C/D snoRNA
CUGA
12345
UUGACAAAACCUGCA
UGCAGGUUUUGUCAA
Box D
RGUAUGA
Box C'
2'-OMe
3'-
5'-
RNA substrate
3'-

Additional observed snoRNA properties. The snoRNA structures in general follow the rule shown in Figure 4, but there are a few exceptions. U85 scaRNA was the first scaRNA discovered as a box C/D and box H/ACA hybrid. This composite scaRNA seems to be functional as overexpression of a snRNA chimera has shown correct 2'-O-methylation and pseudouridylation (Jády and Kiss 2001). Since then, 3 more composite scaRNAs, as well as dual boxC/D and dual box H/ACA snoRNAs have been identified. These special snoRNAs (scaRNAs) were manually curated due to highly altered secondary structures and folding dynamics. Although the number of these snoRNAs are quite low (13), they are fairly conserved across species dating back to early vertebrates (Lestrade 2006, Marz et al. 2011). Therefore, further investigation might provide insights on the importance of conservation in these RNAs.

Although the number of curated known and predicted box C/D snoRNAs in human alone amounts to more than 250 in human and close to 350 in mouse, a significant portion of them have no known methylation targets; conversely, yeast possesses much less snoRNAs and majority of them are well known, most likely due to its inherently small genome (Table 2). Compared to rRNAs, which are abundantly expressed and where matching 2'-O-methylation guide snoRNAs have been identified, finding the targets of the snoRNAs with no obvious complementarity to rRNAs has proven to be difficult (Jády & Kiss 2000). The major hurdle presented to the community has been that each base pairing sequence between a snoRNA guide and its target is so short. Validating predicted pairs is not feasible without already having a list of high confidence modification sites for cross checking, as predictions based on base pairing alone can generate hundreds or even thousands of potential methylation sites. Thus, these snoRNAs without known targets are designated "orphan" snoRNAs (Jorjani et al 2016).

Table 2. Summary of snoRNAs from curated database.

Species	Total # of known and predicted C/D snoRNAs	# with rRNA targets	# with snRNA targets	# with unknown targets
Human	213	129	5	79
Mouse	327	132	5	193
Yeast	47	45	-	2

snoRNA list summarized from snoRNA Orthological Gene Database (snoopy):

http://snoopy.med.miyazaki-u.ac.jp/snorna_db.cgi (Yoshihama & Kenmochi 2013)

Interestingly, a number of snoRNAs of both box C/D and box H/ACA types, and both orphan and non-orphan have also been found to be processed into miRNA like structures and possess miRNA activities. Some of these snoRNA derived miRNAs such as SNORD78-miRNAs have been reported to be correlated with non-small- cell lung cancer, indicating potential functions outside of the conventional scope (Saraiya & Wang 2008, Ender et al. 2008, Scott et al. 2009, Brameier et al.2011). Although it is possible that the effects mentioned are purely contributed from miRNA forms of these snoRNAs, it has not been explored whether potential 2'-O-methylations are carried out by these snoRNAs, whether the methylation levels also correlate with pathology of said cancers. In addition, snoRNAs are found to be part of two types of unusual RNAs: sno-lncRNAs and SPA-ncRNAs (Wu et al. 2016, Yin et al. 2012). Functions claimed in these studies attempt to explain Prader Willi syndrome in a way that complements our hypothesis that 2'-O-methylation plays an important or even central role, and they will be briefly discussed in a later section.

Mechanism of RNA guide-independent 2'-O-methylation. As mentioned above, ribose methylation in eukaryotes mainly employs a snoRNA independent mechanism for tRNAs and miRNAs. For tRNAs, most of the well characterized methyltransferases, termed Trm(s), were discovered in bacteria and yeast. Trms belong to a class of S-adenosyl-L-methionine (AdoMet)-dependent enzymes in the SPOUT superfamily (Tkaczuk et al. 2007 and references therein). Currently, Trm7 has been identified to directly deposit methyl group at 2'-O positions on bases of tRNA-Phe, tRNA-Trp, and tRNA-Leu in yeast. FTSJ1 has since then been reported to be human homolog to the Trm7 (Pintard et al. 2002, Guy et al. 2017, Somme et al. 2014, Bourgeois et al. 2017 and references therein). Hen1, a helix-loop-helix (HLH) family transcription factor harboring both a dsRNA binding domain, as well as a 2'-O-methyltransferase domain, is another

2’-O-methyltransferase ubiquitously expressed (Begley et al. 1992, Yu et al. 2005). Owing to flexibility in its RNA binding domain, Hen1 is able to bind to and catalyze 2’-O-methylation of the 3’- ends of piRNAs and miRNAs, whose precursors exist in partially double stranded forms. Similar to Trm7, Hen1 is highly conserved and homologs are found across species within eukaryotes (Brown et al. 1992, Kirino& Mourelatos 2007, Saito et al. 2007).

In yeast and mammals, 2’-O-methylation of mRNA cap structures is also carried out through non-guide-RNA pathway. mRNA 5’-capping is an important process for mRNA biogenesis. Successful capping protects RNA transcripts from 5’to 3’ exonuclease degradation, enables nuclear export and is required for most translation initiation events in the cytoplasm. A typical 5’-cap structure consists of cap0, cap1 and cap2 components, where the inverted 7-methylaguanisime (cap0) is linked to two cap1-cap2, which are the first and second transcribed bases, through triphosphate linkage. (reviewed in Smietanski et al. 2014 & Hocine et al. 2010). FTSJD1 and FTSJD2, were identified to be the methyltransferases for mRNA cap1 and cap2 through methyltransferase domain homology to tRNA and rRNA 2’-O-methyltransferases (Trm7 and fibrillarin). As part of the final cap structure, virtually all mRNAs are methylated at cap0 and cap1, while only a subset of the mRNAs contains cap2 2’-O-methylation (Furuichi et al. 1975, Werner et al. 2011).

Functions of ribose 2’-O-methylation. The translation machinery is one of the most vital components of biological units across all domains of life. In mammals, mRNA translation is carried out by ribosomes in the cytoplasm with the help of tRNAs in decoding mRNAs. The ribosomal complex is responsible for synthesis of all cellular proteins. Even ‘lifeless’ viruses rely host ribosomal machinery for propagation. Due to the important nature of the translation, the

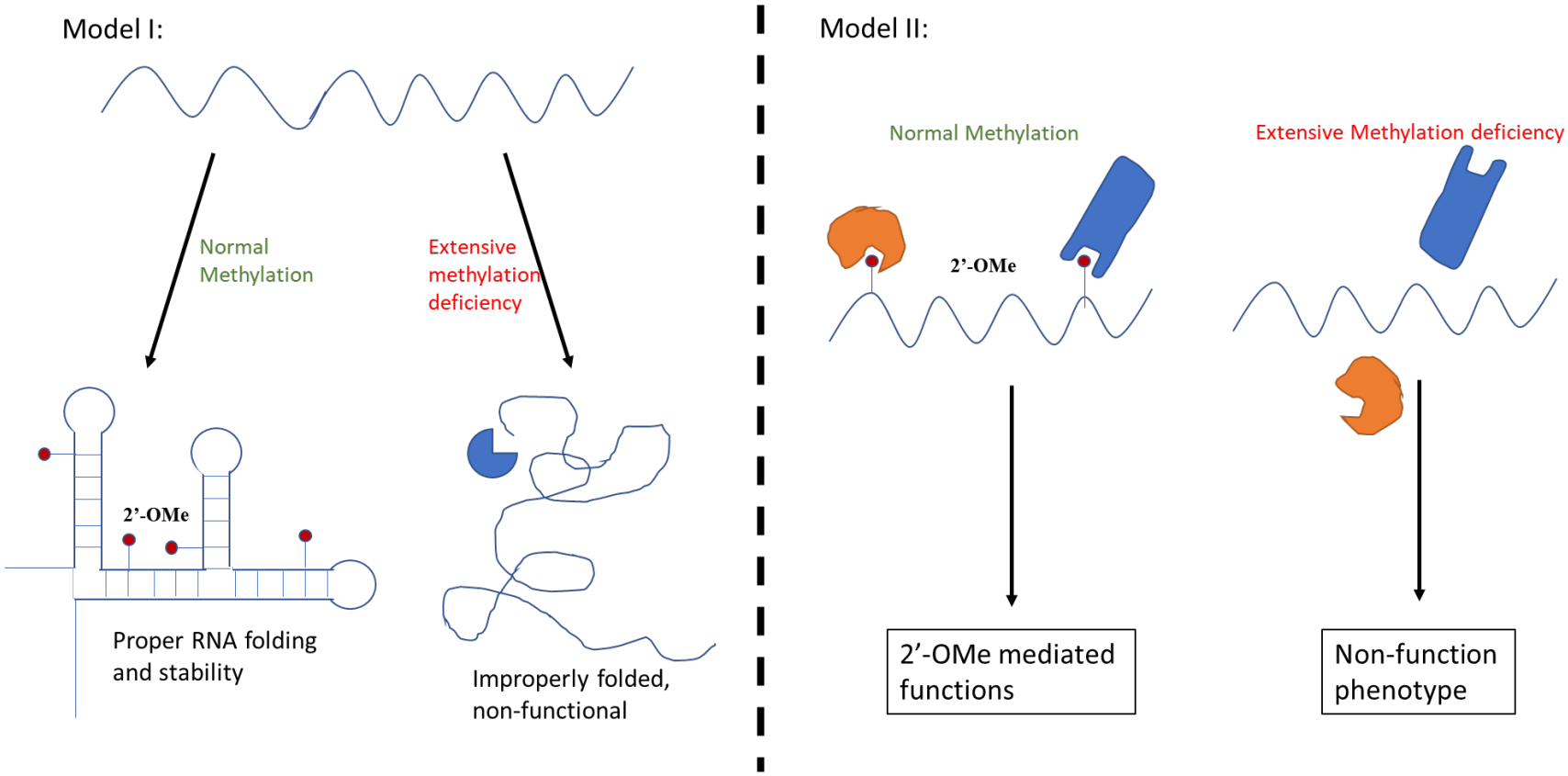
upstream and downstream processes are heavily regulated as evidenced by plethora of transcriptional and post transcriptional regulation pathways. (reviewed in Walsh & Mohr 2011).

As one of best characterized RNA species, human rRNA contains at least 150 2'-O-methylated bases. Although known methylation patterns on rRNA are generally conserved from yeast to human, mostly occurring in regions close to the peptidyl transferase center (PTC), functions of these epitranscriptomic marks remain inconclusive (Decatur & Fournier 2002). Several studies attempting to elucidate exact functions of the modification have resulted in underwhelming results. Serial or aggregated deletions and or mutations of tens of box C/D snoRNA guides that targets methylation sites near core region of rRNA yielded mild phenotypes and growth defects (Esguerra et al. 2008 and references therein). Experiments where multiple modification sites were unmethylated seemed to have produced slightly stronger phenotypes albeit mild nonetheless; similarly, snRNAs, which also possess extensive and conserved 2'-O-methylation patterns, have been shown to have mild defects when snoRNAs targeting them are manipulated (Karijolich & Yu 2007 and references therein). Taken together, these data suggest that 2'-O-methylation on rRNAs and snRNAs might function as structural enhancement features that contribute to RNA transcript rigidity without individually being essential (reviewed in Sloan et al. 2017,). Surprisingly, homozygous knock out of fibrillarin, the 2-O methyltransferase component of the box C/D snoRNPs, proved to be lethal in yeast and resulted in development lethality in mouse embryos; meanwhile, heterozygous KO of the gene had no noticeable phenotype, suggesting a strong tolerance for protein level fluctuations; interestingly the same study showed that correct localizations of the snoRNAs are also disrupted under homozygous KO conditions, consistent with the requirement of the methyltransferase binding for snoRNA maturation (Newton et al. 2003). Although not reported in the publication, there has been no

available fibrillarin KO cell line or conditional fibrillarin KO mouse, which might signify that depletion of fibrillarin is also lethal even later in developmental stages. Such effects, contrary to the putative function of the methylation, suggest that there are two potential models. First, snoRNAs are structurally essential. Depleting fibrillarin depletes snoRNAs and the cumulative effect of elimination of all 2'-O methylation is lethal. Second, at least some specific ribose modification(s) is/are functionally essential, requiring catalytic activity of fibrillarin, and there are site(s) not known associated with lethal effect resulted from fibrillarin knockout (Figure 5). For 2'-O-methylation, although it is highly likely that it contributes to RNA structural stability, as well as influences folding as evidenced by its distribution patterns in different RNA species and studies performed on rRNA biochemical properties (Dennis et al. 2015, Satoh et al. 2000), we cannot disregard the possibility that the modification has additional functions and is more broadly distributed over the transcriptome at the same time.

Figure 5. Models for potential roles of 2'-O-methylation.

See text for details.


Model I:
Normal Methylation
Extensive methylation deficiency
2'-OMe
Proper RNA folding and stability
Improperly folded, non-functional
Model II:
Normal Methylation
2'-OMe
Extensive Methylation deficiency
2'-OMe mediated functions
Non-function phenotype

RNA-independent ribose methylation biochemically achieves the same goal as snoRNA guide methylation in eukaryotes and Archaea, however, their substrates are largely different. The strong conservation of various 2'-O-methylation mechanisms suggests that having the modification is likely conferring some type(s) of evolutionary advantage. It has been known that there exist everlasting power struggles between organisms, such as higher eukaryotes against viruses, throughout the evolutionary process. As viruses evolve, other organisms have to keep up the pace to fend off potential viral infection. Such back and forth competition has shaped our immune mechanisms and viral infection strategies. 2'-OMes at mRNA 5'-cap in higher eukaryotes and some viral transcripts seem to be one hallmark of such a process. Hosts utilize 2'-OMe on the 5'-cap to distinguish self from non-self since mRNA from viruses that replicate in the cytoplasm are modified differently. The lack of 2'-OMe at cap structures can elicit interferon induced responses, thus attenuating viral replication. Interestingly, some viruses encode their own 2'-O-methyltransferases, evading such host immunity (Daffis et al. 2010, Hyde & Diamond 2015, Züst et al. 2011). Further, non-cap internal 2'-OMe was found in falvivirus NS5 transcripts. The presence of this modification reduced the rate of translation and RNA synthesis, which might indicate an alternative strategy to elude host immune system detection (Dong et al. 2012). A similar strategy has also been found to be exploited by bacteria, where a single specific ribose methylation on a bacterial tRNA is enough to antagonize TLR7/8 induced innate immunity (Rimbach et al. 2015, Schmitt et al. 2017). In another unusual case associated with immunity, a naturally occurring 2'-OMe on human 18S rRNA has been reported to be capable of stimulating, not suppressing, the TLR7/8 response pathway (Jung et al. 2015). Although, it is unclear whether this methylation event is linked to any potential function, it might be worth studying under the scope of autoimmunity.

Although current lines of evidence suggest 2'-OMe could be multi-functional, a comprehensive description of the consequences of modification is lacking. Taking all available findings into consideration, it is clear that we need more data points to make proper generalizations. Here are some of the questions we can ask. How strongly do internal 2'-OMes of some viruses affect viral life cycles and how widespread, is this mechanism? Is TLR7/8 activating 2'-OMe only a rare case? Do 2'-OMe functions vary, and by how much, between cell types? How variable is 2'-O-methylation on rRNA, and does 2'-O-methylation variability correlated with cellular phenotypes? Can RNA-guided methylation occur in the cytoplasm or nucleoplasm, instead of only in the nucleolus and Cajal bodies?

C. Insights into 2'-OMe in humans.

Despite the observation that snoRNA derived miRNA-like RNAs are found in the cytoplasm, full-length snoRNAs have never been shown to localize the same way. A notable analogy supporting guide RNA-dependent 2'-O-methylation on mRNA is ADAR family protein localization. From imaging experiments, snoRNPs seem to almost exclusively localize to the nucleolus and Cajal bodies, consistent with their putative functions. Nuclear ADAR1 and ADAR2 also show exclusive nucleolar localization; however, their dsRNA substrates are nucleoplasmic. Indeed recently, a group of ribosomal protein L13a (Rpl13a) intron encoded snoRNAs were discovered in the cytoplasm. While U32a, U33, U34 and U35a transcribed from the locus mainly reside inside typical locations in the nucleolus, stress conditions, specifically reactive oxygen species, triggered significant relocation into cytoplasm in a NADPH oxidase (Nox) regulated manner (Michel et al. 2011). These snoRNAs are guides for known rRNA 2'-O-methylation sites, further expanding the multi-functionality of box C/D snoRNAs. This study showed that the snoRNAs are in the form of intact snoRNPs instead of potentially processed sno-

miRNAs, suggesting any function implicated by the Nox regulated snoRNA relocation is independent of known mechanisms (Holley et al. 2015). Furthermore, the study also detected most of the known snoRNAs in the cytoplasm after ROS triggered stress, therefor identifying a broader spectrum of the phenomenon than with Nox alone.

Confirmation of cytoplasmic snoRNAs opened the further possibility that snoRNPs could indeed modify mRNA internal sites, thus providing one potential explanation to major missing links between snoRNAs and human cancers, which have also been studied extensively recently. Currently, clinical studies have created correlations such as loss of: SNORD50 in prostate cancer, SNORD42 in non-small cell lung cancer, SNORD47/76 in glioblastoma, SNORD47/113 hepatocellular carcinoma and SNORD44 head and neck squamous cell carcinoma and breast cancer, while SNORD50 is deleted in around 10% cases of most human cancers; even a fraction of box C/D snoRNAs are now characterized as biomarkers in leukemia (Patterson et al. 2017 and references therein, Ronchetti et al. 2013, Siprashvili et al. 2013). Could these snoRNAs direct 2'-O-methylation on mRNAs with an effect in cancer? On the other hand, we cannot exclude the possibility that rRNA 2'-O-methylations are heavily involved in these diseases. Contrary to the notion that epigenetic regulation is considered part of transcriptional regulon, which in general has high level of variability between different organ tissues, epitranscriptomic modifications have not been widely explored in a similar way regarding ribosomal activity. SnoRNA expression levels have been known to exhibit distinct patterns under stress and disease conditions (reviewed in Stepanov et al. 2015). It is possible that rRNA 2'-OMe sites are dynamic, rather than static, features that can contribute to a layer of regulation during translation. Owing to potential complex interactions, it is possible that differential methylation of rRNAs might enable regulation of specific genes or subsets of genes. This notion is echoed in Xue &

Barna (2012), who proposed a model of dynamic ribosome biology. More interestingly, further investigation into ribosomes from within a single cell line revealed specialized ribosomal complexes, which have specific ribosomal protein compositions, and are active in translating a subset of messenger RNAs (Shi et al. 2017). It was further corroborated by the new evidence that rRNA sequences between different rRNA clusters are in fact different, rather than just repeats of the same sequences, suggesting potentially different binding affinities for various protein, RNA interaction partners (Kim et al. 2018).

SnoRNA-guided 2'-O-methylation has been shown to be easily reconstituted *in vitro* without needing a long list of factors, suggesting that snoRNPs may be functional wherever they are localized within the cell. An *ex-vivo* study in *Xenopus* oocytes by injection of branch point 2'-O-methylated pre-mRNAs showed strong preference of alternative splicing using cryptic branch points (Ge et al. 2010, Galardi et al. 2002). This report certainly demonstrated possible nucleoplasmic functions of RNA-guided 2'-O-methylation.

The Prader-Willi critical region locus on chromosome 15 q11-q13 represents one of the most complex genomic regions that exhibits extensive splicing, anti-sense transcription and imprinting; this region is associated with the neurodevelopmental diseases Prader-Willi syndrome (PWS) and Angelman syndrome, but the exact molecular basis of the diseases has not been clearly understood (Figure 6, Chamberlain et al. 2010 and references therein). Within this locus are two clusters of orphan box C/D snoRNAs, the SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) clusters (Cavaillé et al. 2000, Bortolin-Cavaillé & Cavaillé 2012, reviewed in Cassidy et al. 2012). SNORD115 was identified as a brain-specific and imprinted snoRNA cluster (45 identical snoRNAs) and was predicted to target the serotonin receptor 5-HT2RC gene through sequence complementarity. This predicted target coincides precisely with

known ADAR editing sites on the serotonin gene. Subsequent studies demonstrated that 1) The SNORD115 snoRNAs form canonical snoRNPs. 2) Expression of SNORD115 inhibits ADAR2 editing of the serotonin receptor. 3) These snoRNAs influence exon inclusion decision on 5-HT2RC (Vitali et al. 2005, Kishore & Stamm 2006). The SNORD116 cluster of 30 closely related snoRNAs is always deleted in PWS patients, but no targets have been identified (Bieth et al. 2015).

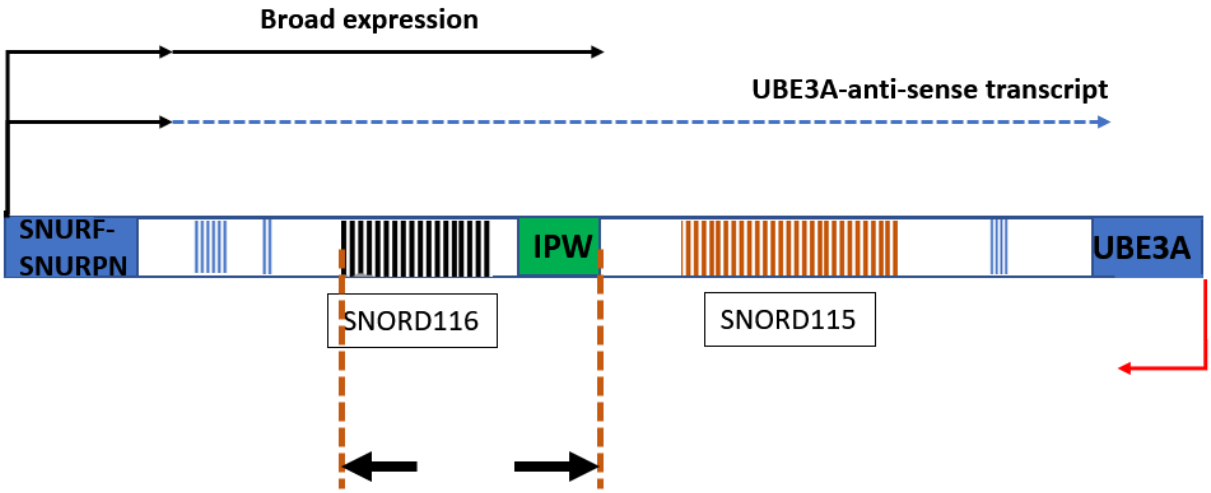
Although no processed sno-miRNAs are found for SNORD116 cluster, two unusual RNA species are transcribed from the region: sno long non-coding RNAs (sno-lncRNAs), where a lncRNA is flanked by two snoRNA ends; and 5'-SnoRNA capped and 3'-PolyAdenylated (SPA) lncRNAs (SPA LncRNAs) (Yin et al. 2012, Wu et al. 2016). Unlike snoRNAs, sno-lncRNAs have predominantly nucleoplasmic localization and contain multiple binding sites for the splicing factor RBFOX2, a protein that is known to regulate alternative splicing. Both sno-lncRNAs and SPA lncRNAs can form nuclear aggregations that trap at least RBFOX2 and two other proteins, and it has been proposed that the two lncRNAs function as sponges to sequester splicing factors and thus disrupt some normal splicing processes (Yin et al. 2012, Wu et al. 2016, reviewed in Li & Fox 2016). Although RBFOX2 overexpression in tissue culture affected the inclusion of a number of alternative exons as reported from sno-lncRNA study, the physiological relevance of the sequestration model remains to be validated. Interestingly, RBFOX2 has recently been reported to have a role in attenuating global Polycomb Complex II activity, suggesting that the sno-lncRNA and SPA-lncRNAs may affect more than splicing (Wei et al. 2016). It is still not resolved whether, SNORD116 snoRNAs can behave similarly to SNORD115 in terms of effects on splicing, or whether SNORD115 or SNORD116 induced 2'-O methylations on specific targets contribute to PWS.

Figure 6. Diagram for the genomic locus of the Prader-Willi Syndrome critical region.

Anti-sense transcription is paternally silenced. Solid arrow represents the broad transcription activity that is common in most cell types. The **dashed blue** arrow represents the UBE3A antisense transcript that is only generated in neurons. UBE3A transcription from the sense strand is maternal allele specific in neurons, while bi-allelic transcription occurs in other cell types. Dashed brown lines indicate the minimum deletion region associated with PWS patients (Bieth et al. 2015).

chr15q11-q13

Imprinted locus, Prader-Willi Syndrome (PWS) critical region



PWS minimum deletion region

So far, I have detailed numerous possible links between 2'-O-methylation and various observations. Although the report of viral mRNA internal 2'-O-methylation is promising (Dong et al. 2012), there remains a significant lack of direct evidence of the existence of 2'-O-methylation on sites/genes of interest other than on rRNAs or snRNA. We can now clearly see the benefit of obtaining the methylation landscape, be it aiding in characterization of orphan snoRNAs, or investigating whether presence of a 2'-OMe correlates with splicing processes or studying methylation variability. It would allow us to study the modification more purposefully instead of relying on the chance of serendipitous discovery of new sites and/or potential functions.

D. Insights into 2'-OMe in other organisms.

Like human, the mouse also expresses a large number of snoRNAs. Consistent with the conservation of snoRNA machinery, not only snoRNP mechanisms are highly similar, introns of mouse snoRNA host genes are highly homologous to those of human counterparts despite much greater variation in the exons of the respective host genes (Ganot et al. 1999, Yoshihama et al. 2013, Tanaka-Fujita 2007). Therefore, rodent samples have been used extensively in conjunction with human and yeast samples for many of the discoveries made in the field. Since mouse samples are more readily available compared to other higher eukaryotes while also being genetically and physiologically closer to human than most other model organisms, it could be an even more powerful tool now for snoRNA and 2'-O-methylation studies as we investigate further into the disease aspect of the snoRNA machinery.

Some distantly related non-higher eukaryotic organisms exhibit particularly high levels of 2'-O-methylation and the significance, besides less extensively studied, of this is also not yet clear. *T. brucei* is a protistan parasite causing deadly disease in sub-Saharan Africa – the African

sleeping sickness-. During its life cycle the parasite shuttles between its tsetse fly vector and its human host where it lives freely in the bloodstream (reviewed in Ponte-Sucre 2016). Trypanosomes possess 21 snoRNA clusters that are responsible for producing more than 90 snoRNAs, of which 57 are box C/D. It was suggested by bioinformatic mapping and partial experimental validations that there are more than 130 2'-OMe rRNA sites in *Trypanosome brucei* (Uliel et al. 2004, Liang et al. 2005). Interestingly, differential methylation on the rRNA has been reported on two different life cycle stages of the organism. Consistent with the earlier mentioned dynamic ribosome model, the differential methylation detected was not a change in the overall level, but rather, changes in specific sets of sites (Barth et al. 2008). Differential pseudouridylation has recently also been implicated in this transition (Chikne et al. 2016). The authors have inferred from thermophilic organisms, many of which have high numbers of 2'-O-methylation sites, that the differential pattern could be a way help coping with increased temperature going from insect form to mammalian blood stream form. However, this assumption may rather predict an overall increase in methylation of the rRNA. Further investigation is required to understand the precise function of the 2'-OMe in *T. brucei*. Systematic profiling of the sites could be a starting point, followed by biophysical and other studies to determining their structural contributions, as well as whether they are involved in additional functions. Since *T. brucei* is a rather remote organism from higher eukaryotes in the evolutionary tree, studying it would also aid in further comprehension of the robustness of the regulatory roles of the 2'-OMe in human.

E. Methods to study 2'-OMe.

2'-OMe sites were traditionally detected using one of four strategies. The first strategy employs primer extension. The primer extension experiment utilizes polymerase pausing effects

when obstacles are encountered during extension. Such an effect is amplified when the experiment is carried out under limited dNTP concentrations, especially when encountering secondary structures and some bulky modifications. If RNA gel electrophoresis is performed on these primer extension products, bands corresponding to sites of polymerase stoppage can be seen. Thus, to detect/validate a suspected site, a downstream primer can be designed, and a primer extension experiment can be carried out (Figure 7A, Maden et al. 1995). A second method is based on the fact that ribose methylated bases are resistant to alkaline hydrolysis. If an RNA pool, such as total RNA, is hydrolyzed to shorter fragments, it is expected for any methylated RNA that hydrolysis will occur at every position except for the site of the methylated base. RT can then be performed with a primer specific to a transcript suspected to have 2'-OMe. The RT product is then visualized on a gel together with a sequencing ladder. For example, if only a base 150bp upstream of the primer is 2'-O-methylated, a fragment of 150bp would not exist after partial alkaline hydrolysis, and this would correlate with a gap at that base on a sequencing gel (Figure 7B, Kiss-László et al. 1996). A third method was developed based on the observation that RNaseH cleavage strictly requires perfect pairing between DNA and RNA molecules. For a suspected 2'-OMe site on human 18S rRNA for example, a short DNA oligo can be designed to hybridize to the region. After RNaseH digestion, visualization on a gel should show either presence or absence of the digestion products depending on whether the methylation indeed occurs at the region (Figure 7C, Yu et al. 1997). A final method is to purify a subset of RNA transcripts that are of interest and subject them through liquid chromatography and mass spectrometry after fragmentation into proper sizes (Qiu & McCloskey 1999).

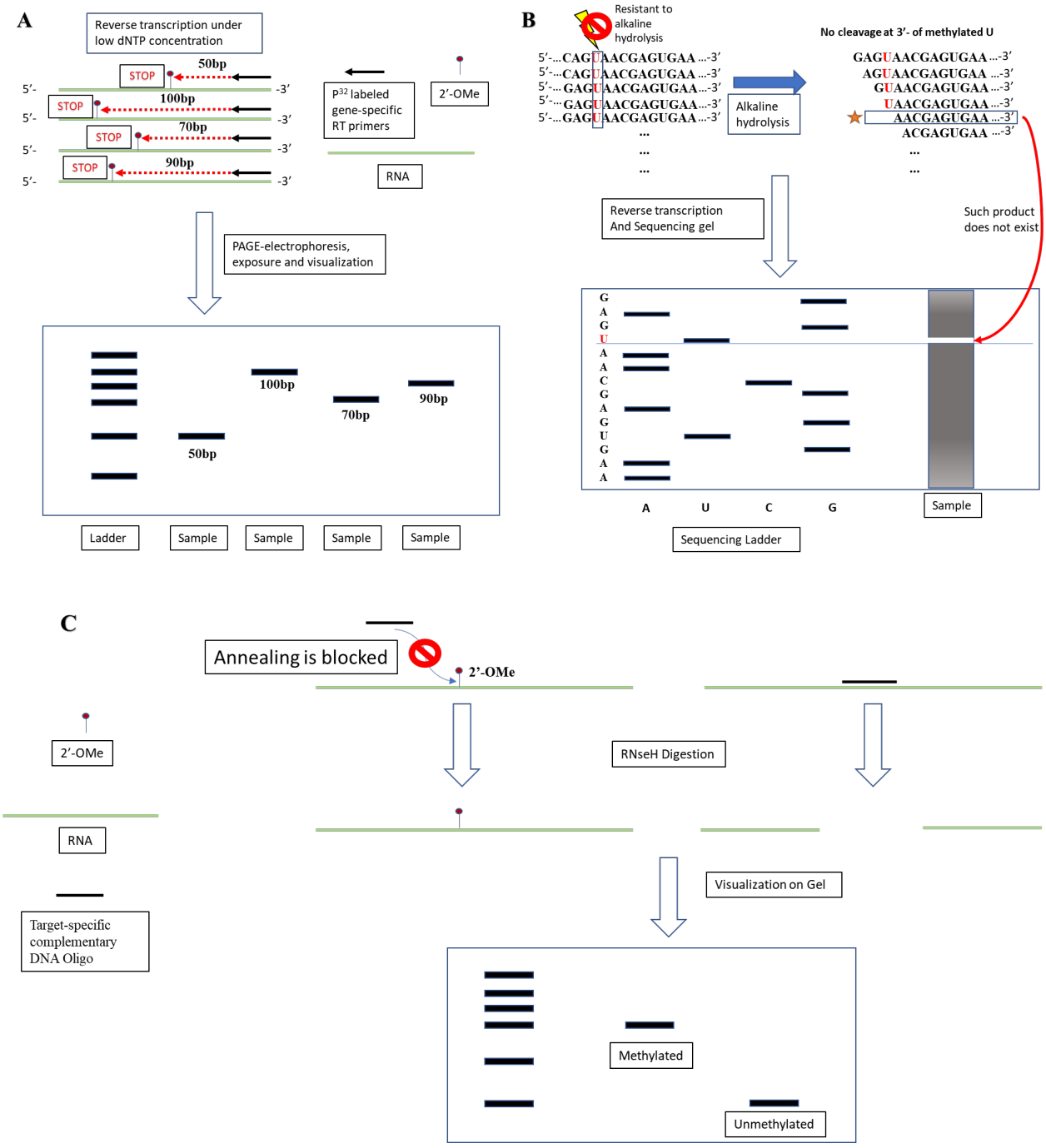
Figure 7. Illustrations of conventional methods used for 2'-OMe detections and validation.

(A) Low dNTP Primer Extension. Each RNA diagram represents different transcripts with different 2'-OMe sites. Gene-specific primers are designed for targets of interests. Under low dNTP concentration, 2'-OMe is enough to induce significant amount of pausing of the polymerase extension, resulting in a product pool with major species as the paused fragments.

(B) Alkaline Hydrolysis. Sequences in the diagram are multiple copies of one RNA transcript form of a hypothetical gene. Red-color highlighted bases are 2'-O-methylated. Because 2'-O-methylated bases are resistant to alkaline hydrolysis, the digestion cannot occur at 3'- of the 2'-O-methylated U base (the star labeled sequence). Thus, if the random digested product is reverse transcribed with a gene-specific primer, and visualized on a sequencing gel, there will be signal for every base that corresponds to hydrolyzed RNA fragments shown in top right panel. However, there will be a gap at 2'-O-methylated U position as a fragment from hydrolyzing the U cannot be generated.

(C) RNaseH Assay. 2'-OMe disrupts perfect oligo-RNA annealing, thus preventing RNaseH digestion from happening. To test whether a several-base region contains a 2'-OMe, a radiolabeled 10bp oligo targeting the specific loci can be designed, and RNaseH assay carried out. The digestion product can be visualized on PAGE gel.

Methods Used in Detecting sites of 2'-O-methylation.



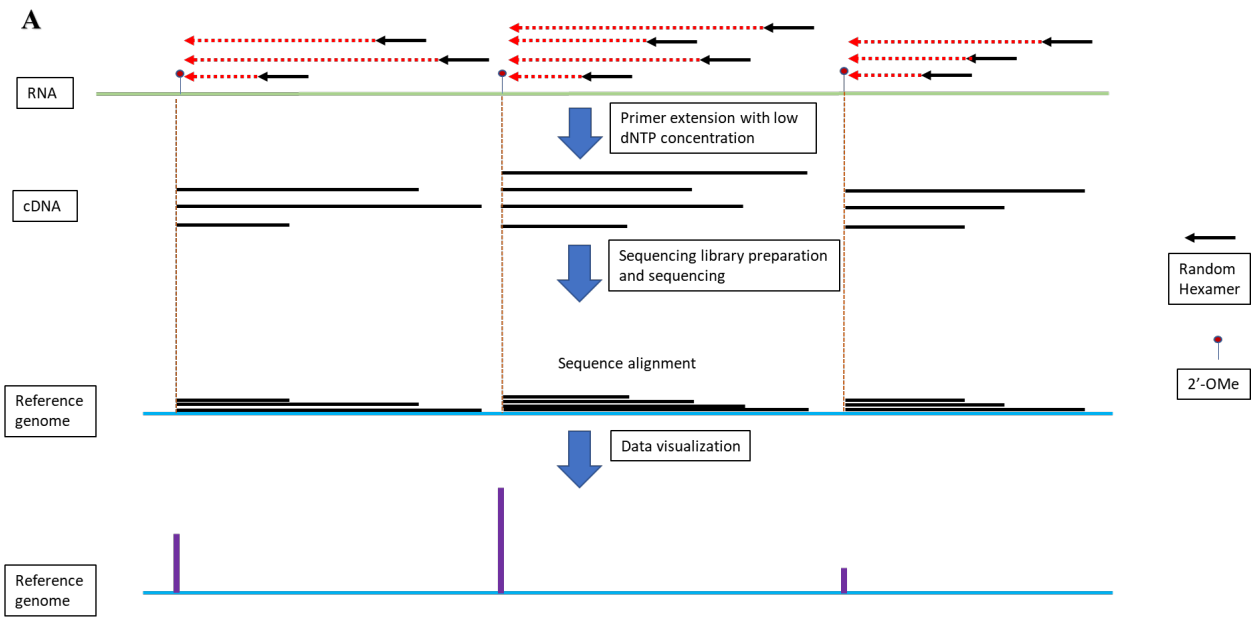
Although multiple methods were available for detecting a single modification, major drawbacks exist. In primer extension under low dNTP conditions, despite observations of polymerase pausing, 2'-OMe is a much weaker factor compared to RNA secondary structures and some other modifications that strongly affect base pairing (reviewed in Maden 2001). Such properties make this method highly prone to false positives for moderately long transcripts. It could be viable to pinpoint to a small stretch of RNA bases, but even then, one has to be aware of local secondary structures and other modifications. The RNaseH method is limited as a general confirmation tool due to its inability to map at single base resolution besides the same factors plaguing primer extension method also causing uncertainties of any results. The biophysical method using mass spectrometry remains highly accurate, however, it is also the most labor and cost intensive method and is unsuitable for the discovery of new sites. Furthermore, all methods described lack scalability. Although these methods were invaluable in identifying majority of 2'-OMe sites on rRNAs, the less than 150 sites total detected were the result of decades long work and piecing together data from various labs. These methods at their current state are acceptable for validation purposes, but not suitable for explorative studies.

To address the shortcomings of the conventional methods, two high-throughput strategies have been developed in recent years. RIM-seq and 2OMe-seq modify the low dNTP primer extension method and adapt it to Illumina sequencing workflow, thus enabling high -throughput profiling (Jorjani et al. 2016, Incarnato et al. 2016). Unlike primer extension experiments mentioned above, RIM-seq and 2OMe-seq use random primers, removing the restriction to the analysis of single transcripts; the products of primer extension are converted into Illumina sequencing libraries, sequenced, and analyzed (Figure 8A). However, the disadvantage of non-specific pausing remains in this approach. In RIM-seq, more than 400 2'-OMe sites were

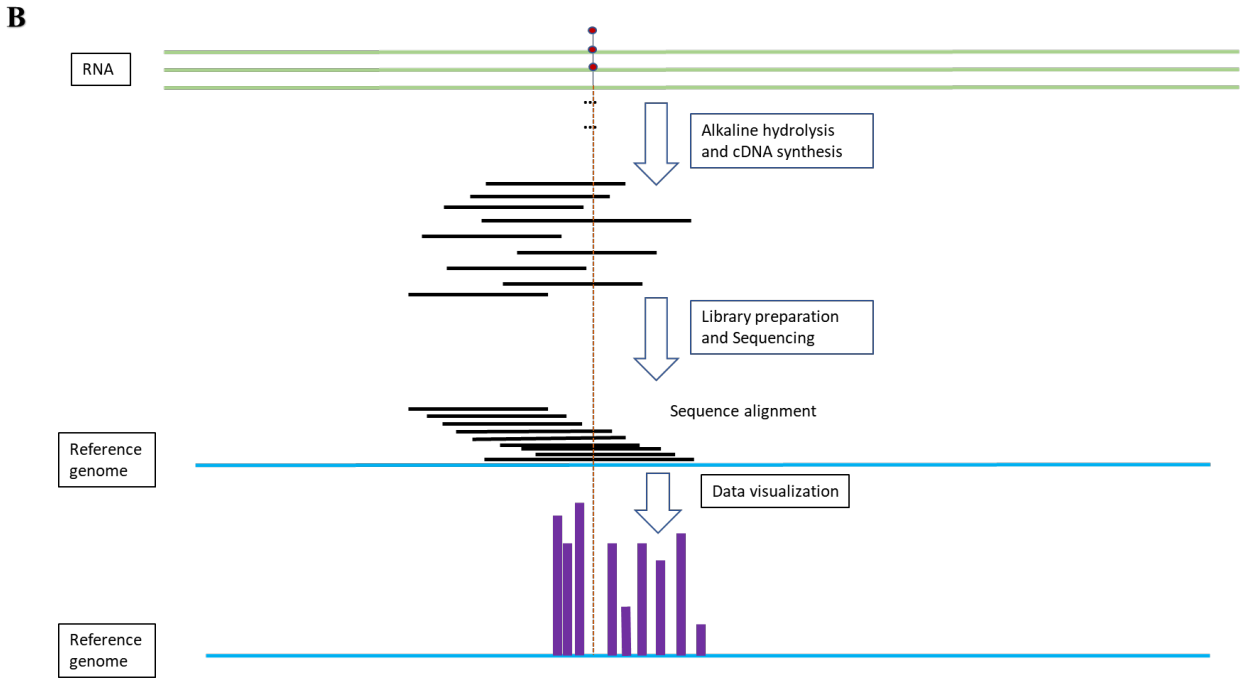
identified, at least 200 more than previously known, but it is unclear how many of the novel sites are true positives owing to the inherent high false positive rate of primer extension. 2OMe-seq attempts to partially alleviate the issue with the addition of a parallel fibrillarin knockdown experiment. The KD samples' mapping should show decrease of signal for sites that are snoRNA guided. Thus, any sites detected in the non-KD samples that correspond to reduction in KD samples theoretically represent true sites. However, besides the necessity to double the sample amount, gene silencing experiments are known to introduce variables (Olejniczak et al. 2009). Ribometh-seq, on the other hand, combines alkaline hydrolysis with NGS. In this case, the alkaline hydrolyzed RNAs are directly converted into sequenceable cDNA libraries and sequenced (Marchand et al. 2016). The premise is that due to randomness of the hydrolysis, when enough RNA is supplied, there should be hydrolysis at every single position of all RNAs except for methylated bases. Thus, the presentation of the alignment would show mapped 3'-ends on every single base position, while gaps would form on methylated ones (Figure 8B). Although the method has the advantage of being highly specific, its reliance on negative rather than positive signals can become an issue when applied to lower abundance RNAs such mRNAs. In addition, it is viable for rRNAs as there are less than 8kb distance to cover. However, for human mRNA with average length of about 3kb and around 100000 different types transcripts, a normal exome RNA-seq requires 30 million reads per sample, translating to 0.09 read per base, while Ribometh-seq would require 300 reads per-base, or 100 billion reads per sample for accurate 2'-OMe site calling. The cost for such sequencing effort is unattainable for most laboratories.

Figure 8. Available high-throughput methods for 2'-OMe detection.

(A) RIM-Seq/2OMe-seq. Primer extension with low dNTP concentration is carried out with random primers. The RT cDNA product pool is then converted into a sequencing library. After sequencing and read alignment, read 5'-end enrichment is calculated for every base position. **(B) Ribometh-seq**. RNA is randomly hydrolyzed under alkaline condition. The digested fragments are converted into sequencing library and sequenced at very high depth. Because 2'-O-methylated-base is resistant to hydrolysis, there does not exist any fragment that ends with that methylated base position. Thus, when read 3'-end enrichment is calculated for every base position, 2'-OMe site corresponds to the base position lacking read 3'-end mapping.



Peaks correspond to 2'-OMe sites



A lack of peak corresponds to 2'-OMe site

F. Thesis Objectives.

As biotechnologies leaped forward in the past decade, the technological barriers preventing us from accurately and efficiently profile 2'-OMe sites are melting away. As discussed above, even with recent developments, our ability to accurately and efficiently profile 2'-OMe sites still remains unsatisfactory. To be able to test the 2'-OMe function centric model, and to answer the questions we proposed so far, we need a flexible and efficient way of profiling the sites transcriptome-wide. The main focus of this thesis, thus, will be the development of a highly specific and accurate high-throughput method detecting 2'-OMe sites. In Chapter II, I will summarize the strategy we used to selectively enrich 2'-OMe sites and its distinct features comparing to currently available methods. I will also present the developmental timeline and optimization steps taken. Finally, tests of these principles on human rRNA sites will be discussed. In Chapter III, I will discuss the current state of the method, as well as applications of the method. Due to the highly flexible nature of the method, I will be emphasizing on different directions we can take to improve and adapt the method in Chapter VI.

Chapter II

Development of Ribose Oxidation Sequencing (Riboxi-Seq) For Profiling 2'-Ome Sites

A. Abstract

RNA modifications, or epitranscriptomic marks, are not only important for basic biological processes such as nominal ribosomal functions, but also have been implicated in cellular defects and diseases. Ribose methylation (2'-O-methylation, 2'-OMe) occurs at high frequencies in rRNAs and other small RNAs and is carried out using at least one shared mechanism across eukaryotes and archaea. It is one of the epitranscriptomic marks that have been identified to be involved in many cellular processes without clearly characterized functions. Evidence so far indicates 2'-OMe being mainly a structural feature, however, due to deficiencies in methodology studying the modification, there might be more to be learned even on rRNAs. Furthermore, the possibility of internal 2'-OMe sites on mRNA has been suggested (Dong et al. 2012, Holley et al. 2015), thus it is important to characterize the landscape of 2'-O-methylation. Here we report the development of a highly sensitive and flexible method for ribose methylation detection using next-generation sequencing. A key distinction of this method is to use sodium periodate to enrich 2'-OMe sites, allowing only RNAs harboring 2'-OMe groups at their 3'-ends to be sequenced. Although currently requiring microgram amounts of starting material, this method is robust for the analysis of rRNAs even at low sequencing depth. An important feature of the method is that it can be expanded to work on a transcriptome-wide scale with minimal revision.

B. Background

A great majority of 2'-O-methylations are directed by Box C/D snoRNAs, noncoding RNAs that guide the modification of target sites via complementary RNA sequences. In humans,

snoRNAs are assembled into snoRNP particles, containing the conserved core proteins NOP56, NOP58, 15.5K, and fibrillarin (the catalytic component) (Tycowski et al. 1996; Filipowicz & Pogači 2002; Watkins & Bohnsack 2012). 2'-O-methylation has been extensively studied for a number of years with the goal of establishing functional and mechanistic links with specific biological pathways. Although, early studies demonstrated that 2'-O-methylations on rRNAs are prevalent during ribosome biogenesis, depleting several snoRNA guides alone or generating fibrillarin heterozygous mutations does not produce noticeable defects (Newton et al. 2003). However, homozygous mutations of fibrillarin have been shown to be embryonically lethal in mouse, thus suggesting that this gene is essential for organism development (Tollervey et al. 1993, Newton et al. 2003). Such evidence strongly suggests that the modification could have important roles. 2'-O-methylation has also been shown to be present on tRNAs and has been implicated to be crucial in translational circuitries (Satoh et al. 2000; Guy et al. 2015). A substantial portion of known methylated sites in rRNA lie in close proximity to ribosome functional sites such as regions around the peptidyl transferase center, suggesting the potential involvement of such modifications in rRNA folding, stability, and translation (Decatur & Fournier 2002). Ribose methylated bases are also found at mRNA caps and are involved in host pathogen responses (Daffis et al. 2010; Rimbach et al. 2015). Additionally, 2'-OMe adenosines have been found within, rather than at the cap, of flavivirus RNA transcripts. Homology between 2'-O-methyltransferases of Flavivirus, Dengue-1 and West Nile also suggest a more common phenomenon (Dong et al. 2012). Furthermore, recent evidence indicates that in addition to being associated with the 5' cap, mRNAs might potentially possess internal 2'-O-methylated sites (Lee et al. 2016, Holley et al. 2015). Thus, currently available data hint that there is an increasing

possibility that 2'-OMe possesses much more expansive regulatory roles than in ribosome biogenesis.

The list of known 2'-O-methylation sites is frequently updated, as experimental techniques evolve and mature. However, until recently, a major hurdle in obtaining a more complete profile of the 2'-O-methylation landscapes has been the lack of an efficient and reliable modification-specific and high-throughput detection method. Methylation sites have traditionally been mapped using targeted approaches including primer extension under limiting dNTP concentrations, where reverse transcriptase stalls when encountering a methylation site, or resistance to RNaseH digestion when synthesized DNA oligos are introduced (Yu et al. 1997; Maden 2001). Primer extension experiments are particularly prone to false positives for detecting 2'-O-methyl sites due to nonspecific polymerase pausing or secondary structure-induced pausing, while RNaseH assay lacks base resolution. Most importantly, both methods require laborious mass spectrometry to validate a detected site (Qiu & McCloskey 1999). Also, neither is suitable for de novo site detection and high-throughput screening, because the base position needs to be known in advance for primer or hybridization oligo design. This makes primer extension and RNaseH assay most useful as confirmation tools. Recently, primer extension, as described in RIM-seq and 2OMe-seq, has been adapted for high-throughput detection of ribose methylation sites by combining random priming with next-generation sequencing (Incarnato et al. 2016; Jorjani et al. 2016). This study identified over 400 sites, almost 300 more than what have been curated in human rRNAs (Lestrade & Weber 2006). It is unclear, however, how many of the novel sites are true positives, owing to an inherent high false-positive rate of primer extension. Although potential matches to BoxC/D snoRNAs were bioinformatically identified for some of the novel sites found in the study, methylation could not be confirmed, because there are known snoRNAs

that interact with targets without guiding the deposition of methyl on ribose (Cavaillé & Bachellerie 1998; Lafontaine 2015). As a consequence, these sites might not accurately represent the methylation pattern until they are further validated.

Alkaline hydrolysis is another conventional method used to identify 2'-OMe. Due to the resistance to hydrolysis of a base modified by 2'-OMe, fragmented RNAs can be reverse transcribed and gaps correlating to the sites are visualized on gel (Kiss-László et al. 1996). Although alkaline hydrolysis is very specific to 2'-OMe, the site needs to be known in advance. Ribometh-seq was developed recently and utilizes the property of alkaline hydrolysis of ribose methylated bases. Thus, by randomly hydrolyzing RNA and performing next-generation sequencing at very high depth, there should be uniform coverage of 3'-end positions across regions of interest except at positions of 2'-O-methylation. This method overall has much better specificity and accuracy, as it has successfully detected about the same number of sites in rRNAs as have been annotated. Several novel sites were also validated by mass spectrometry (Krogh et al. 2016). However, the method relies heavily on negative rather than positive signals. In addition, the requirement for high read depth and coverage makes such studies costly, and the method can also suffer from high background noise due to resistance to alkaline hydrolysis of highly structured regions. In order to address these issues, we have developed a 2'-O-methyl ribose-specific, high-throughput method, which relies on positive rather than negative signals, to detect 2'-O-methylation sites.

C. Method development, Methods and Materials

1. Methodology development

Development of an accurate method for 2'-OMe detection is best achieved by utilizing a modification specific strategy. Unlike other modifications such as m^6A and m^1A, there is no

antibody available for 2'-OMe, nor does there exist any 2'-OMe specific chemical modifier that enables purification. As a consequence, other strategies for enriching 2'-OMe had to be considered.

Random digestion is a common strategy for Illumina based regular RNA sequencing, as it reduces RNA transcript to a more uniform length appropriate for the sequencer. The result is that there will be RNA fragments generated such that each fragment represents 5'- and 3'- ends of a portion of the original RNA transcript. When mapping these fragments back to the genomic sequence, a pattern representing distinct 5'- and 3'- ends, each with similar frequency, can be obtained. Such a property is useful because when combined with more stringent digestion, which generates shorter fragments, there will be a higher probability that a random base will be present at 5'- or 3'- end of a fragment (Figure 9A). Ribometh-seq for example exploits this particular chemistry together with ribonucleic acids' resistance to hydrolysis if 2'-O-methylated, thus producing a sequencing pattern where appearances of negative signal are dictated by probability.

However, to achieve a positive signal, resistance to hydrolysis is not enough, because 2'-O-methylated ends can never exist in the digestion product due to their resistance. The closest that random degradation can achieve is 1 base upstream of the 3'-end of a fragment (Figure 9A). To ensure that 2'-OMe sites are represented at the ends of the fragments at all, it is necessary to remove at least 1 nucleotide either from the 3'-end. Removing more than 1 base, which will be discussed later, can greatly increase sensitivity, although it can make the procedure rather laborious. Through literature searches, we found a stepwise strategy to degrade RNA base by base that has been available since the 20th century. In this strategy Alefelder et al. (1998) first oxidized RNA 3'-ends using sodium periodate ($NaIO_4$), which have 2'- and 3'- OH groups (cis-diols), into 2'-, 3'- dialdehydes. Such dialdehyde structures are prone to β-elimination, which

removes the oxidized RNA terminal oligonucleotides. The resulting products will contain terminal 2'-O-methylated bases with 3'-phosphates and nonmethylated bases with a mixture of cyclic-phosphates, 3'-phosphates, and 2'-phosphates (Figure 9B).

By extensively hydrolyzing RNAs to sizes smaller than those obtained in the Ribometh-seq protocol, we speculated that we should be able to generate a pool of RNA fragments that has sufficient 2'-O-methylated ends to provide grounds for the rest of the procedures. However, instead of -OH groups at 3'-terminal bases' 2'-ribose positions required by oxidation, alkaline hydrolysis produces 2'- phosphate groups. To simplify our pilot experiments, we found a commercially available enzyme called Benzonase nuclease, which is both an endonuclease and an exonuclease, and which generates cis-diols when used for digestion (Figure 9C, Sigma-Aldrich). Benzonase nuclease, unlike most other RNAses, does not have a sequence bias. Therefore, we expected digestions with Benzonase to behave similarly to alkaline hydrolysis.

With 2'-OMe exposed, we next sought to enrich these ends. Conveniently, 2'-O-methylated bases have been shown to be resistant to $NaIO_4$ oxidation we used for beta elimination. This could serve as a perfect strategy to convert all non-methylated ends into dialdehydes while keeping 3'-terminal bases with 2'-OMe intact. Indeed, an earlier publication on a quick ligation protocol used this property to convert un-ligated products, which have cis-diols at their 3'-terminal ends, into dialdehydes to enrich ligated products that have unreactive 3'-ends (Kirino & Mourelatos 2007, Kurata et al. 2003). The oxidized RNA fragments should then be ready for sequencing library preparation, which requires ligation of a linker to the fragments' 3'-ends in order for subsequent conversion into cDNA using RT. Because fragments with dialdehydes at their 3'-termini are incompatible with ligation, only the ones that have terminal 2'-OMe groups are converted into the cDNA library. PCR amplification can then be performed to generate the

final library that is labeled with sequences recognized by the Illumina sequencer, which was introduced with the PCR primers (Figure 9D).

18S rRNA (SSU) and 28S rRNA (LSU) add up to about 7kb in length, where there are around roughly 150 known 2'-OMe sites, averaging around 1 2'-OMe per 50 bases. Depending on the efficiency of 2'-OMe site exposure after fragmentation and beta elimination, we expect the final terminal 2'-OMe numbers to be relatively low. As a consequence, one caveat of the proposed library preparation strategy is that after oxidation, the number of fragments available for ligation and conversion into sequenceable reads might become too small without having to drastically increasing starting materials. Two potential issues can occur: first, extremely small quantities in general are hard to quantify; second, PCR preferential amplification can greatly alter original ratios between different fragments. The first problem can be addressed by increasing PCR cycles, however, this worsens the PCR bias outcome. To remedy these issues, we include a random sequence portion into the RT primer. In general, digestion of RNA does not generate fragments with exactly the same 5' and 3' ends. Thus, if two sequenced reads have exactly the same RNA portion and same random sequence, these are most likely generated from PCR amplification, and are thus duplicates. The inclusion of the random sequences allows us to remove these PCR duplicates during data processing, thus reducing the overall bias of the procedure while keeping the starting material amount relatively low (Figure 10A).

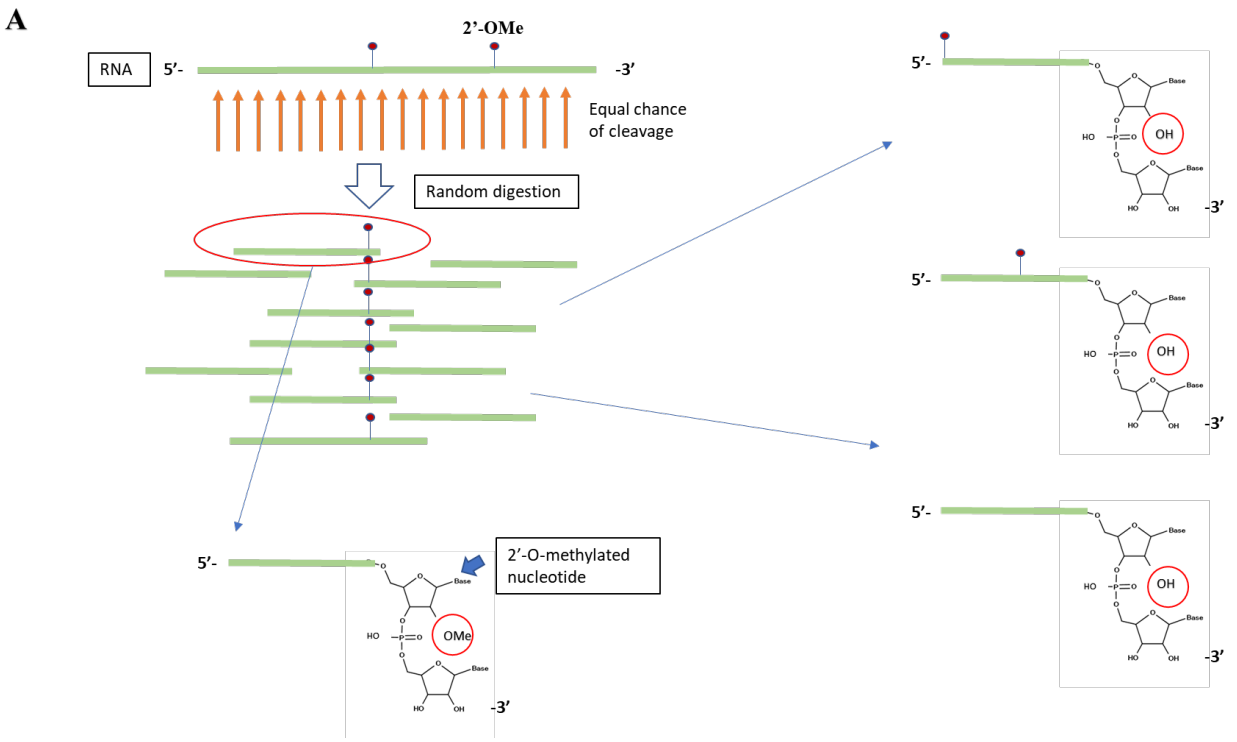
Figure 9. Chemical principle of the RibOxi-seq.

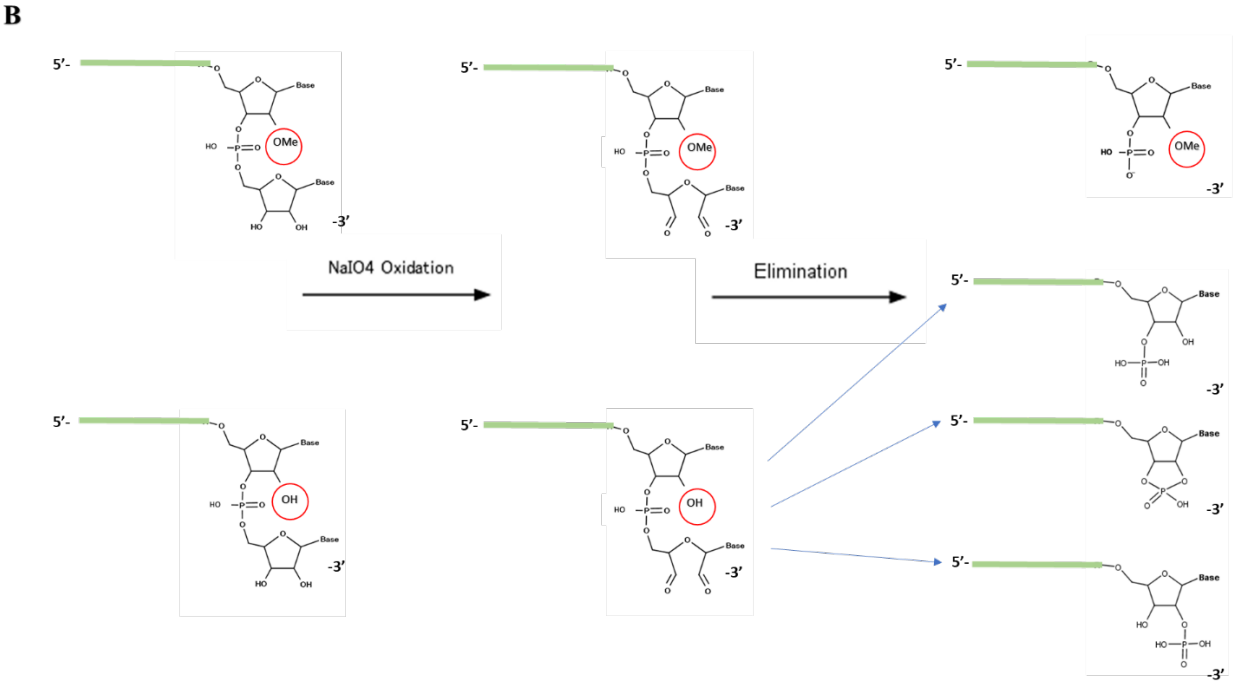
(C) Property of random digestion. In theory, every base position of an RNA transcript can be cleaved, generating fragments with variable 3'-ends. These 3'-ends cannot contain 2'-O-methylated bases as explained earlier in the thesis. The bottom left panel represents the type digestion products where a 2'-OMe is 1 nucleotide upstream of RNA 3'-end, which is the species concerned in the method. The right panel shows other possible background RNA species.

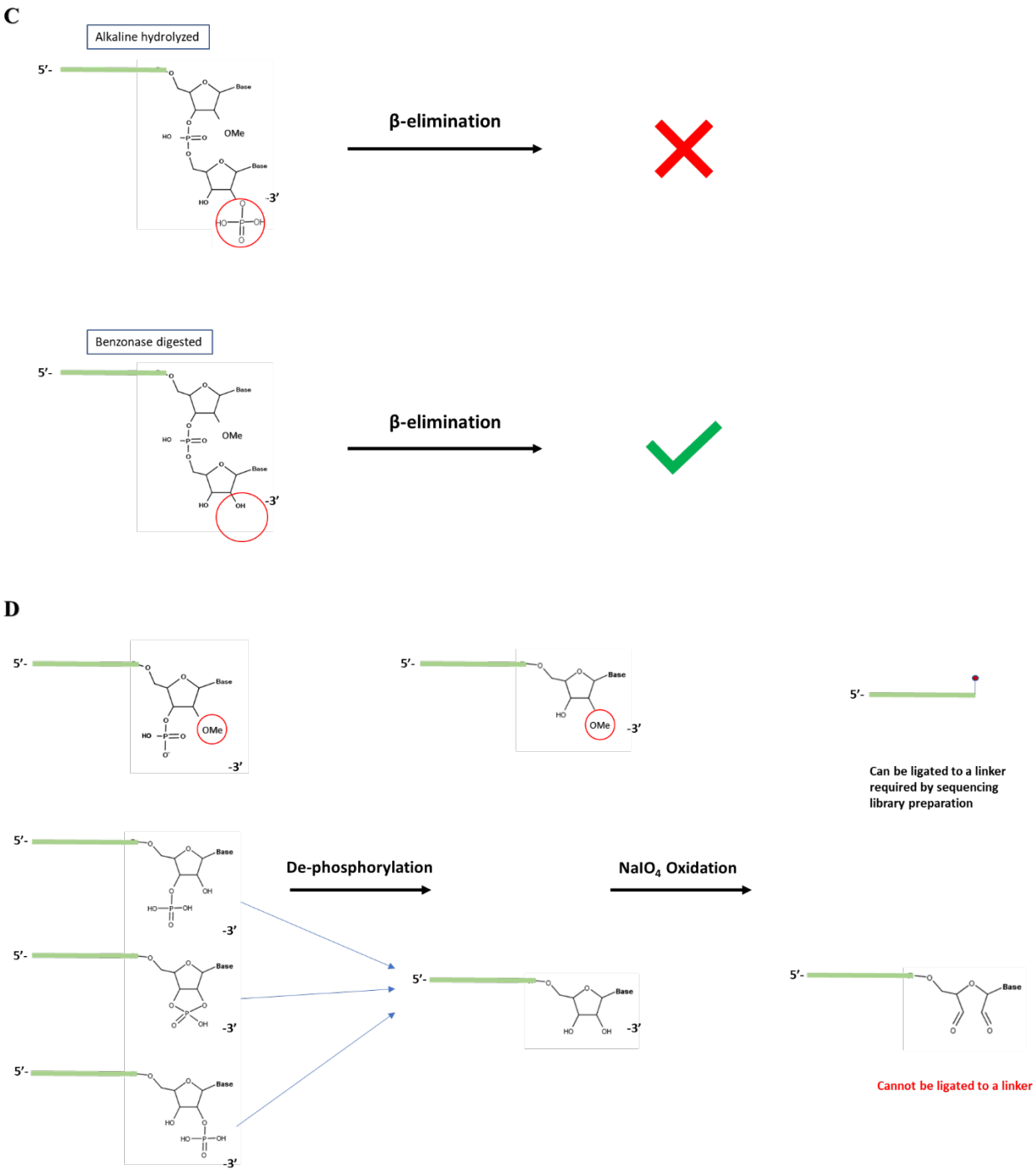
(D) β-elimination. Elimination of one 3'-terminal nucleotide for exposure of 2'-OMe.

(E) Random Digestion strategies. RNA transcripts hydrolyzed under alkaline condition generate fragments with 3'-phosphates, which require extra steps before elimination can be performed. On the contrary, Benzonase digested fragments possess the *cis-diol* structure that is directly compatible with elimination.

(F) Enrichment of terminally methylated RNA fragments. RNA mixture obtained from β-elimination is undergone $NaIO_4$ oxidation a second time. Because fragments with terminal 2'-OMe are protected from oxidation, only non-methylated fragments are converted into di-aldehyde ends. Subsequently, only fragments protected by 2'-OMe can be converted into sequencing library and sequenced.


A
RNA
5′-
-3′
2′-OMe
Equal chance of cleavage
Random digestion
2′-O-methylated nucleotide
OH
OMe
Base


B
5′-
-3′
OMe
NaIO4 Oxidation
Elimination
OH
Base


C
Alkaline hydrolyzed
5′-
Base
OMe
HO
-3′
β-elimination
Benzonase digested
5′-
Base
OMe
HO
OH
-3′
β-elimination
D
5′-
Base
OMe
-3′
5′-
Base
HO
OMe
-3′
5′-
Can be ligated to a linker required by sequencing library preparation
5′-
Base
OH
-3′
5′-
Base
-3′
5′-
Base
HO
-3′
De-phosphorylation
NaIO4 Oxidation
5′-
Base
HO
OH
-3′
5′-
Base
Cannot be ligated to a linker

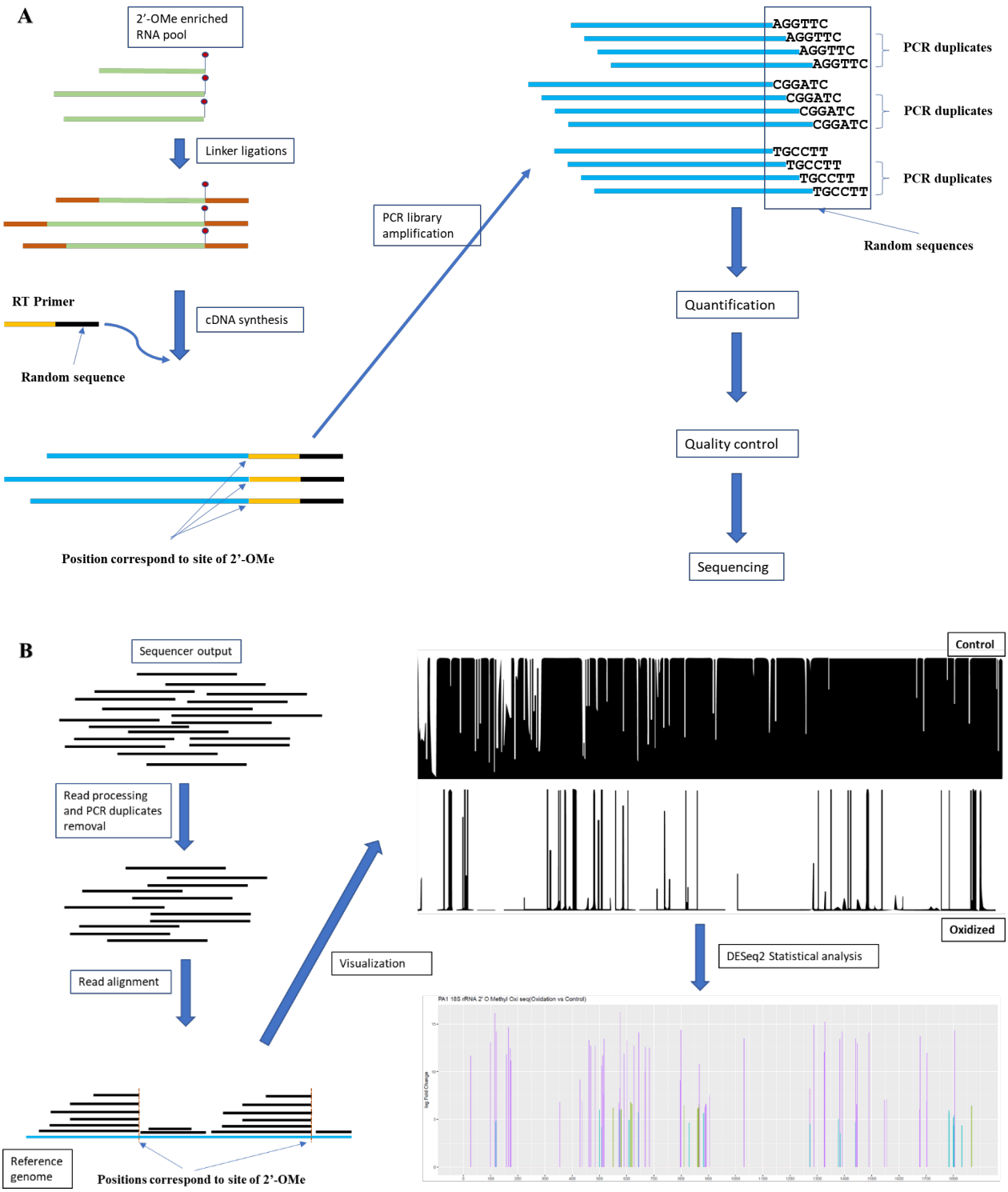
The sequencing reads obtained from the Illumina sequencer first need to be analyzed to remove PCR duplicates discussed above followed by trimming to remove the random sequences, as well as the linker sequences introduced during ligation. The processed reads can then be aligned to a reference genome. The 3'-terminal bases of the aligned reads correspond to the original RNA 2'-O-methylated 3'-ends. We speculated that by counting these ends, we would be able to see a pattern that is largely consistent with the distribution of known 2'-OMe sites (Figure 10B).

In order to statistically evaluate the results, a control group for each sample is needed. To properly control for the method, we proposed to have additional samples undergo exactly the same procedures except for the oxidation step. Such preparations of the control samples should not show enrichment of 2'-O-methylated ends, thus generating a relatively uniform alignment pattern across the reference genome. On the contrary, they should show a lack of ends mapping to 2'-OMe sites reminiscent to the principles of Ribometh-seq (Figure 8B, Figure 10B). Subsequently, end-counting data can be analyzed using DESeq2, a bioinformatic tool for differential expression analysis. DESeq2 compares control data, where ends are distributed evenly, with oxidized data, where ends align at specific positions. The program uses a negative binomial statistical model to evaluate the differential between the two and assigns significance values, as well as computing $\log_2$ fold changes for every position (Figure 10B, Love et al. 2014). We would then be able to examine data and determine the optimum cutoff for 2'-OMe site-calling.

Figure 10. Continuation of RibOxi-seq

(A) Library preparation and PCR duplicates removal strategy. Random hexamer sequence is introduced as part of the RT primer. After PCR amplification, any identical ends together with identical random sequence is considered PCR duplicate and can be removed later bioinformatically.

(B) RibOxi-seq data processing, analysis and visualization. Enrichment of read 3'-ends at sites of 2'-O-methylation (bottom left panel). When counting just 3'-ends, the data can be visualized in top right panel.


A
2'-OMe enriched RNA pool
Linker ligations
RT Primer
cDNA synthesis
Random sequence
Position correspond to site of 2'-OMe
PCR library amplification
AGGTTC
AGGTTC
AGGTTC
AGGTTC
PCR duplicates
CGGATC
CGGATC
CGGATC
CGGATC
PCR duplicates
TGCCTT
TGCCTT
TGCCTT
TGCCTT
PCR duplicates
Random sequences
Quantification
Quality control
Sequencing
B
Sequencer output
Read processing and PCR duplicates removal
Read alignment
Reference genome
Positions correspond to site of 2'-OMe
Visualization
Control
Oxidized
DESeq2 Statistical analysis

To summarize the key principles of the proposed method: first, fragmented RNAs containing 3'-ends that are either unmethylated or 2'-O-methylated are obtained using random digestion followed by β-elimination; then, an oxidation step renders the nonmethylated ends incapable of ligation to linkers used for high-throughput library construction; after sequencing, the reads are aligned to a reference genome and only positions of the 3'-ends of aligned fragments are counted and displayed for each base position; the count data for oxidized and nonoxidized samples are then normalized, compared, and analyzed using DESeq2 for single-base resolution methylation site determination. The major difference between our method and currently available methods is its specificity and its reliance on positive rather than negative signals. The cis-diols of non-2'-O-methylated ribose are converted into dialdehydes using $NaIO_4$, thus preventing them from being ligated to linkers for sequencing library construction. The advantage of this step over alkaline hydrolysis is that Benzonase leaves terminal ends with 2'-OH instead of phosphate group, thus eliminating an extra dephosphorylation step.

Before starting the pilot experiments, it would be to our advantage to determine whether Benzonase nuclease is able to cleave at 2'-O-methylated bases, removing the requirement of eliminating a 3'-terminal base. To test this, we designed an oligo that is fully methylated. A digestion of the oligo shows that Benzonase does not cut at 2'-O-methylated bases (Figure 11A). Therefore, β-elimination is still required.

We started our experiment by optimizing Benzonase nuclease digestion through time course and temperature gradients with PA1 cell total RNA. Benzonase rapidly degrades RNA within minutes at room temperature, leaving little room for optimization. Thus, all digestions were carried out on ice. The aim was to obtain the smallest fragments without exceeding the lower detection limit of the Agilent TapeStation (25bp), which is an imaging system that allows very

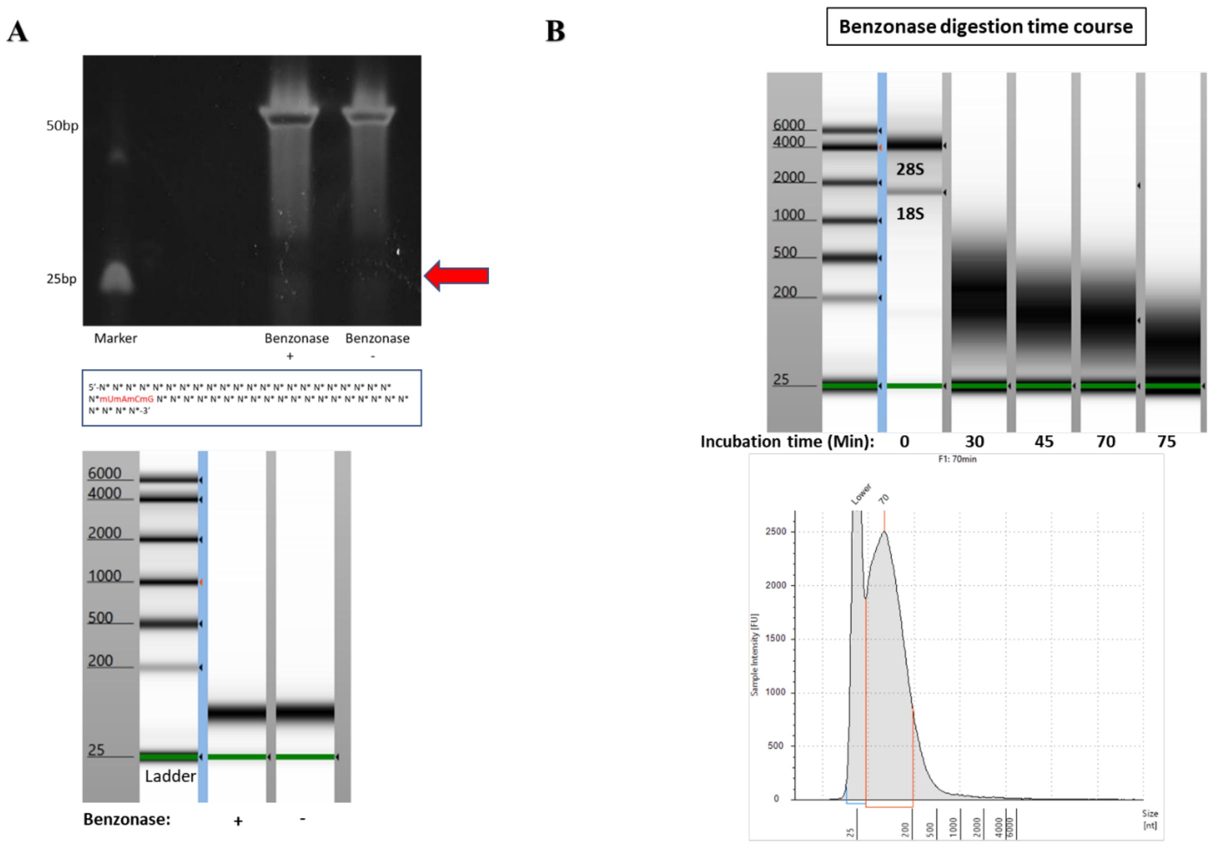
fast visualization and quantification of nucleic acids. Several digestion time courses helped to determine our target sizes to be an average of 70bp (Figure 11B).

Next, we proceeded with β-elimination. As mentioned earlier, the reaction leaves a mixture of terminal cyclic-phosphates, 3'-phosphates, and 2'-phosphates on non-methylated fragments. To enrich methylated ends, oxidation is required to inactivate non-methylated ends, which require terminal cis-diols. Thus, a phosphatase treatment was necessary. Most available phosphatases, such as calf intestine alkaline phosphatase, are only active in removing 3'-phosphates. If used, nonmethylated bases with 2'-phosphates would survive the oxidation step and generate false-positive signals in sequencing. To prevent this phenomenon, T4 polynucleotide kinase (T4 PNK) was used. Although less efficient than other widely used phosphatases, the advantage is that this enzyme nevertheless is capable of removing all three types of phosphates under acidic conditions in the absence of ATP (Cameron & Uhlenbeck 1977, Das and Shuman 2013, Kirino & Mourelatos 2007). It has been shown that T4 PNK's phosphatase activity is sufficient to remove the majority of the phosphates in <40 min at 37°C (Honda et al. 2016). Since RNA fragments with phosphates can generate bias in subsequent steps, we have increased enzyme concentration and extended incubation time to 4 h. With proper end treatment, the mixtures generated after β-elimination and T4 PNK treatment contain fragments with ends of either cis-diol or 2'-OMe,3'-OH.

Figure 11. Protocol optimization

(A) Confirming Benzonase inability to cleave at site of 2'-OMe. 50bp RNA oligo is designed to have locked DNA bases except for the 4 2'-O-methylated RNA bases in the middle. Digestion using Benzonase was visualized on Agarose gel (top) and tapestation (bottom).

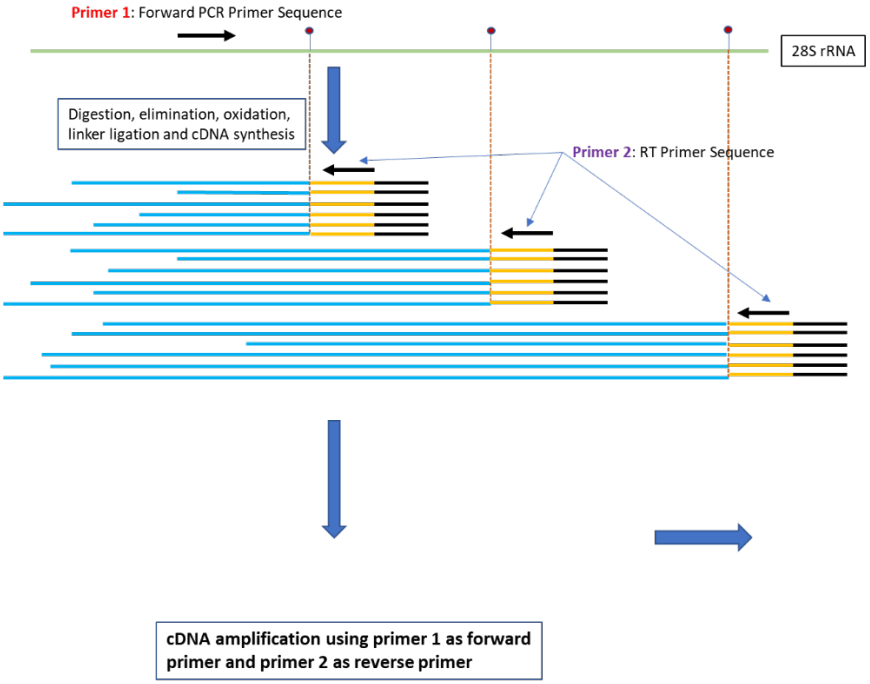
(B) Digestion time course. Bottom panel represents the size distribution corresponding to the last lane in top panel with 75min digestion time.


A
50bp
25bp
Marker
Benzonase +
Benzonase -
5'-N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N*
N*mUmAmCmG N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N* N*
N* N* N* N*-3'
6000
4000
2000
1000
500
200
25
Ladder
Benzonase: + -
B
Benzonase digestion time course
6000
4000
2000
1000
500
200
25
28S
18S
Incubation time (Min): 0 30 45 70 75
F1: 70min
Lower
70
Sample Intensity [FU]
2500
2000
1500
1000
500
0
25
200
500
1000
2000
4000
6000
Size [nt]

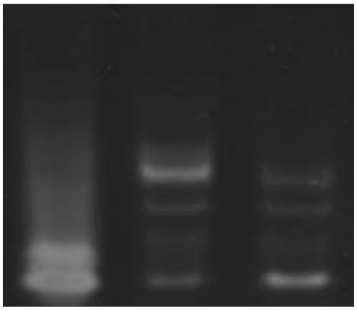
After another $NaIO_4$ oxidation treatment, we needed to ligate a linker to provide sequence complementarity for subsequent reverse transcription primers. Different variants of the T4 RNA ligases are known to have different reactivities toward different substrates and some of them generate bias toward certain base identities. Our goal was to minimize sequence bias during the ligation step and also maximize efficiency. It has been shown that both T4 RNA ligase I and II are biased against 2'-O-methylated bases, while ligase I has further bias against 2'-OMe-U (REF). Under optimized condition however, T4 RNA ligase II is able to ligate pre-adenylated linkers to RNA with 2'-OMe ends at acceptable efficiency, while displaying almost no base identity bias (Munafó & Robb 2010, Zhuang et al. 2012, Raabe et al. 2014). We thus followed the available optimized protocol to ligate 3'- linker. A 5'-linker is then ligated prior to reverse transcription for PCR primer annealing in a later step. We then tested both 5' and 3' ligation using primers complementary to the ligated linkers and showed that ligation was successful. At this step, we wanted to have at least partial confirmation that the basic chemistry works before finishing the library preparation and proceeding to sequencing. We designed 3 rRNA forward primers at distinct locations and PCR amplified the cDNAs of both control and oxidized samples with RT primer as reverse primer. We observed a continuous smear in controls while oxidized samples displayed distinct bands, consistent with the notion that fragments should have a uniform distribution in control but ends accumulating only on methylated positions in oxidized sample (Figure 12). The validation encouraged us to proceed with final steps of library preparation and subsequent sequencing.

Figure 12. Library validation using PCR

Refer to main text for details



cDNA amplification using primer 1 as forward primer and primer 2 as reverse primer



Control Non-oxidized | Oxidized Sample1 | Oxidized Sample2

Visualization on agarose gel

The initial sequencing was done with 2 biological replicates for each control and oxidation experiments on Illumina MiSeq sequencer, which generated a minimum number of reads. Although low read counts may negatively affect accuracy, this was enough for a proof-of-principle experiment. Sequencing results were analyzed with a custom bioinformatic pipeline, which is presented at the end of this Chapter. Comparison of our data with available 2'-OMe databases was performed, to ensure our goal was achieved, prior to increasing to 3-replicates per sample and switching to the Illumina NextSeq platform, which generated 5 times more data.

2. Cell culture and RNA extraction

RNAs used in the initial experiments described in this chapter were extracted from the human ovarian cancer PA1 cell line and HEK 293T cell line. Both cell lines were cultured using regular DMEM 1X media supplement with 10% FBS as well as 1x Penicillin-Streptomycin in 15cm Petri dishes. Cells were lysed on the dish when reaching confluency with lysis buffer from PureLink® RNA Mini Kit followed by RNA extraction and on-column DNase treatment according to the manufacturer's protocol. Total RNAs used in the WT vs. KO experiments were extracted from either wild type, U32 snoRNA KO or U32+U51 snoRNA double KO 293T provided by the Holley lab at Duke University.

3. PCR validation of cDNA

Forward primers for 3 distinct 18S rRNA locations were used in conjunction with RT primer to amplify 1 µL of cDNA obtained from reverse transcription during the RibOxi-seq protocol. Additional 11.5 µL H_2O and 12.5 µL NEB OneTaq master mix was used for the reaction per manufacturer's protocol. Extension time was set to 10 seconds due to the short nature of the templates. The PCR product was run on a 1.5% agarose gel for 2 hrs.

4. Anti-sense oligo (ASO) knock down of U63 snoRNA

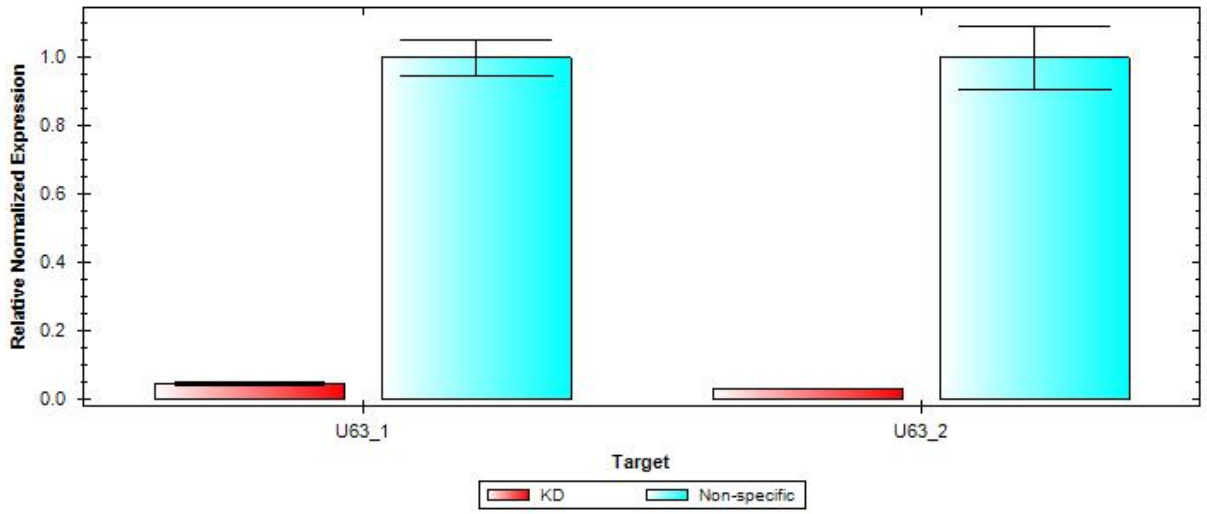
ASOs targeting 2 regions of the snoRNA were designed and ordered from IDT. PA1 cells were cultured as described above. When cultures became confluent, KD was performed using Lipofectamine 2000 following the manufacturer's protocol. Sample setup was: 2 biological replicates for both ASO1, ASO2 and scrambled KD. RNA extractions were performed 48hr post transfection using PureLink RNA mini kit. Knock down efficiency was evaluated by qPCR using miScript II RT Kit and miScript miRNA PCR array kit (Figure 13). RNA was then used in RibOxi-seq once KD had been confirmed.

5. Radioactive primer extension

For novel site validation, primers were ^{32}P labeled. One microgram of total RNA, 1 μL (10 μM) labeled RT primer, 1 μL 100 μM or 1 μL 10 mM (control) concentrations of dNTPs, 7 μL of water were denatured at 65°C for 5 min, then chilled on ice. An RT Master mix (10 μL per reaction) containing 2 μL 10× RT Buffer, 1 μL (40 U) RNase Out, 1 μL AMV RT (NEB), and 6 μL water was prepared, and added to the RNA/primer mix. Incubation was at 42°C for 45 min. Reactions were ethanol precipitated and resuspended in loading buffer for TBE-PAGE electrophoresis. This experiment was performed for three selected sites from RibOxi-seq: positive control at 28S C1880 (lanes 1,2), known but not detected 18S U1668 (lanes 4,5) and newly detected and not previously reported 28S A3717 (lanes 6,7). The first lane of each set is a negative control where primer extension was performed with a higher dNTP concentration. The second lane of each set was performed at low dNTP concentration to promote polymerase pausing at sites of 2'-OMe.

Figure 13. QPCR verification of U63 KD efficiency.

U63_1 and U63_2 represent two U63 specific ASO used for KD. Non-specific KD was carried out using scrambled ASO.


Relative Normalized Expression
1.0
0.8
0.6
0.4
0.2
0.0
U63_1
U63_2
Target
KD
Non-specific

D. Results

1. RibOxi-seq accurately identifies annotated 2'-O-methylation sites within 18S and 28S rRNAs.

RibOxi-seq was used to analyze total RNA from the human ovary teratoma-derived PA1 cell line. Site detection was filtered by a combination of log2 fold change and adjusted P-value from the DESeq2 output. All site-annotations and numbering correspond to the hg19 reference genome. The lists of known sites we used were curated as previously described (Krogh et al. 2016). By applying a cutoff value of log2 fold change of >7 and adjusted P-value of <0.0001, 39 out of 40 known 18S sites and 60 out of 66 known 28S sites were detected with high confidence (Table 3, Table 4). The filters were set to correspond to the known sites that have the lowest log2 fold changes and the highest P-values to allow maximum sensitivity (Figure 14). The number of high confidence sites consisted of 93.3% known sites, which include sites newly found and MS validated by Krogh et al. (2016) using Ribometh-seq. Using such cutoffs, only three novel sites (18S: U354, 28S A1322, and A3717) were found. However, when filters were slightly relaxed to log2 fold change of >6 and adjusted P-value remained unchanged, eight total potential novel sites were identified (Table 4). Among these candidates, A3717, which displays both a very high log2 fold change (∼10) and a low adjusted P-value (∼1.5 × 10−13), in 28S was validated using a primer extension under restricted dNTP concentration (Figure 15). Blasting A3717 in combination with surrounding bases within the snoRNA database resulted in two potential snoRNA guide hits, HBII-180B and U37, whose guides are complementary to this region and fulfill criteria for methylation at the fourth or fifth base of the rRNA complementary sequence. Although these two snoRNAs were previously noted to also methylate other positions within 28S, snoRNA-guided methylations of multiple targets is known to occur (van Nues et al. 2011).

As only a few novel sites of methylation in rRNAs in human have been characterized within recent years, this result argues that the RibOxi-seq method not only identifies known 2'-OMe sites, but also allows the discovery of new ones.

Table 3. List of 2'-O methylation sites found in 18S and 28S rRNA from PA1 cells

18S:

chrUn_gl000220 coordinate	Base position	Base	snoRNAs	Novel?	Detected?	log2FC	adjusted p value
109104	27	A	U27	no	yes	17.92943	3.43E-36
109176	99	A	U57	no	yes	12.66704	1.68E-191
109193	116	U	U42A/B	no	yes	12.43167	2.18E-69
109198	121	U	mgh18S-121/Z17B	no	yes	14.57216	3.9E-249
109236	159	A	U45A/C	no	yes	9.515039	6.76E-08
109243	166	A	U44	no	yes	12.78353	3.81E-12
109249	172	U	U45A/B	no	yes	10.19354	1.51E-11
109251	174	C	SNORD45C	no	yes	11.79013	5.83E-73
109505	428	U	HBII-202	no	yes	9.514961	5.07E-69
109513	436	G	HBII-429	no	yes	7.66968	1.64E-45
109539	462	C	U14A/B	no	yes	12.0917	9.85E-34
109545	468	A	SNORD83A/68	no	yes	12.55015	4.02E-50
109561	484	A	U16	no	yes	11.59956	1.64E-25
109586	509	G	HBII-95/B	no	yes	7.985352	3.07E-10
109589	512	A	HBII-234	no	yes	12.83362	1.98E-74
109594	517	C	U56	no	yes	14.91788	2.85E-71
109653	576	A	HBII-336	no	yes	13.86214	9.17E-81
109667	590	A	U62A/B	no	yes	10.33406	6.34E-10
109679	601	G	HBII-251/U103/B	no	yes	12.56987	1.75E-22
109704	627	U	HBII-135	no	yes	9.429529	3.18E-12
109721	644	G	U54	no	yes	12.67708	9.15E-18
109745	668	A	U36A/B	no	yes	11.75379	3.81E-16
109760	683	G	HBII-108/B	no	yes	14.08851	2.06E-51
109874	797	C	ZL107/GGgCD20	no	yes	9.685468	2.90E-63
109876	799	U	U105/B	no	yes	10.46362	5.47E-06
109944	867	G	HBII-419	no	yes	10.21481	2.95E-46
110108	1031	A	U59A/B	no	yes	7.713374	2.73E-05

110349	1272	C	HBII-142	no	yes	7.509314	7.58E-14
110365	1288	U	HBII-55	no	yes	12.14759	4.76E-68
110403	1326	U	U33	no	yes	8.592145	4.66E-07
110405	1328	G	U32A	no	yes	8.730437	1.33E-06
110460	1383	A	SNORD30	no	yes	11.30044	1.58E-19
110468	1391	C	U28	no	yes	14.38748	2.14E-42
110519	1442	U	U61	no	yes	8.874819	3.31E-11
110524	1447	G	SNORD127	no	yes	11.02978	3.65E-119
110567	1490	G	U25	no	yes	9.173948	4.75E-09
110745	1668	U	Unknown	no	no		
110755	1678	A	U82	no	yes	12.47255	5.01E-51
110780	1703	C	U43	no	yes	11.16467	9.52E-40
110881	1804	U	U20	no	yes	13.96577	2.07E-79

28S:

chrUn_gl000220 coordinate	Base position	Base	snoRNAs	Novel?	Detected?	log2FC	adjusted p value
113745	397	A	U26	no	yes	8.330468	5.02E-34
113747	399	A	U81	no	yes	12.62798	5.39E-41
114663	1315	G	U21	no	yes	9.655468	2.27E-52
114670	1322	A		yes	yes	7.772448014	1.46E-46
114673	1325	A	U18A/B/C	no	yes	11.56306	1.42E-40
114687	1339	C	U104	no	yes	8.451424	8.58E-20
114869	1521	G	snR39B	no	no		
114871	1523	A	U32A/B/U51	no	yes	10.86955	1.06E-20
114881	1533	A	U77/U80	no	yes	8.162373	1.46E-07
114972	1624	G	U80	no	yes	11.99951	1.80E-38
115107	1759	G		no	yes	9.31923	1.37E-25
115218	1870	A	U38A/B	no	yes	10.96471	4.15E-14

115228	1880	C		no	yes	10.55289	7.79E-34
115639	2291	C	U48	no	no		
115698	2350	C	U24	no	yes	9.587046	8.89E-23
115710	2362	A	U76	no	no		
115712	2364	C	U24	no	yes	11.3781	2.84E-77
115748	2400	A	HBII-202	no	yes	9.007025	2.18E-26
115762	2414	U	ZL5/6/SNORD143/144	no	yes	11.36035	2.97E-23
115769	2421	C	mgh28S-2409	no	yes	13.18647	6.38E-171
115771	2423	G	mgh28S-2411	no	yes	10.01012	3.45E-17
116134	2786	A	HBII-420	no	yes	7.649203	6.31E-05
116151	2803	C	U55	no	yes	8.838489	1.27E-79
116162	2814	A	U95	no	yes	12.29032	7.88E-34
116171	2823	C	U95	no	yes	10.82754	1.51E-60
116184	2836	U	U34	no	yes	11.68902	7.29E-20
116208	2860	C	U50	no	yes	11.05747	9.81E-28
116223	2875	G	U50	no	yes	9.26956	1.08E-19
116974	3626	G		no	yes	11.41354	1.91E-47
117049	3700	C	HBII-180A/B/C	no	yes	11.8941	1.28E-35
117065	3717	A	HBII-180B	yes	yes	10.24109708	1.54E-13
117067	3719	A	U37	no	no		
117071	3723	A	U36C	no	yes	10.50822	2.07E-14
117091	3743	G	HBII-276	no	yes	6.466373	5.50E-09
117107	3759	A	U46	no	yes	8.50116	4.32E-10
117132	3784	A	U15A/B	no	yes	10.15055	8.93E-11
117139	3791	G	SNORD15A	no	yes	10.0732	1.37E-29
117155	3807	C	mgU6-77	no	yes	6.963681	3.84E-23
117165	3817	U	ACA48/HBI-43	no	yes	12.21898	1.79E-45
117172	3824	A	U30	no	yes	11.44786	5.37E-33
117177	3829	A	U79	no	yes	13.18303	1.53E-54
117188	3840	C	U74	no	yes	13.59508	9.18E-208

117214	3866	A	HBII-316	no	yes	10.11454	4.60E-95
117216	3868	C	U53	no	yes	11.55245	3.30E-129
117234	3886	C	U47	no	yes	9.185954	5.57E-10
117246	3898	G	HBII-99/B	no	yes	10.27247	1.36E-30
117272	3924	U	U52	no	yes	9.274559	1.51E-11
117291	3943	G	HBII-82/B	no	yes	12.25001	1.73E-127
117391	4043	G	U102	no	no		
117401	4053	C	U75	no	yes	8.835343	1.99E-05
117543	4195	G	U31	no	yes	12.39906	5.00E-31
117574	4226	U	U58C	no	yes	9.428109	4.73E-32
117575	4227	G	U58A/B/C	no	yes	10.08335	1.34E-49
117653	4305	U	U41	no	yes	11.59332	8.11E-19
117717	4369	G	U60	no	yes	9.848911	1.48E-13
117739	4391	G	snR38A/B/C	no	yes	10.2434	6.89E-19
117803	4455	C	U49A/B	no	yes	11.95609	2.11E-135
117841	4493	G	HBII-210	no	yes	10.71001	3.04E-57
117845	4497	U	SNORD62A/B	no	no		
117846	4498	G	SNORD62A/B	no	yes	10.83552	5.26E-85
117870	4522	A	U29	no	yes	9.798517	1.09E-12
117883	4535	C	U35A/B	no	yes	12.49586	9.45E-251
117918	4570	A	U63	no	yes	8.764135	7.07E-11
117937	4589	A	SNORD119	no	yes	11.60263	8.32E-131
117965	4617	G	HBII-296A/B	no	yes	8.407214	1.45E-52
117967	4619	U	HBII-240	no	yes	10.46333	1.10E-90
117970	4622	G	U78	no	yes	11.90306	4.03E-95
117984	4636	G	SNORD121A/B	no	yes	9.767133	1.42E-39

Sites determined by filtering DESeq2 analysis output (libraries sequenced on Illumina nextseq) using log2FC (>6) and adjusted p value (0.0001). Positions highlighted in blue are sites that are annotated but not detected using our method.

Table 4. **List of potential novel sites found in PA1 cells using RibOxi-seq.**

chrUn position	Base position	base	snoRNA	Novel?	Detected?	Log2 fold change	Adjusted p value
18S:							
109431	354	U		yes	yes	8.162207	2.42E-07
109981	904	A		yes	yes	6.712134	1.57E-13
110632	1555	U		yes	yes	6.122584	4.72E-18
28S:							
114670	1322	A		yes	yes	7.772448	1.46E-46
114671	1323	A		yes	yes	6.73175	3.05E-10
114672	1324	A		yes	yes	6.638669	1.08E-09
117065	3717	A	HBII-180B/U37?	yes	yes	10.2411	1.54E-13
117233	3885	G		yes	yes	6.083689	2.33E-21

18S: U354, 28S: A1322 and A3717 were detected even with stringent filter (log2 fold change > 7 and adjusted p value < 0.0001), while the rest were found with analysis threshold set to log2 fold change > 6 and adjusted p value < 0.0001.

2. RibOxi-seq results confirm methylation heterogeneity within the same cell line.

Among already annotated 2'-O-methylation sites, U4497 and G4498 in 28S are positioned immediately adjacent to each other. One limitation of the RibOxi-seq method is that if two sites back to back are both methylated, the site closer to the 5'-end cannot be detected if the other site is fully methylated. This is because Benzonase (as well as other ribonucleases and alkali) cannot cleave at the 3'-site of 2'-OMe, so the 5'-site can never be exposed using our approach. Thus, if G4498 were fully methylated, U4497 would not be detected. Indeed, we did not see U4497, indicating G4498 may be fully methylated. However, another set of back-to-back pairs of 28S sites (U4226 and G4227) were both detected with high confidence. The most probable explanation for our results is that G4227 is only partially methylated, with the unmethylated population allowing the exposure of U4226 (Table 4). This result is consistent with the data obtained using Ribometh-seq as well as observations of fractional methylation from primer extension experiments (Maden 1986; Krogh et al. 2016). Such patterns prompted us to consider the possibility that annotated 2'-O-methylation sites not detected by our method may be the result of a complete lack of methylation. To test this possibility, radioactive primer extension with low dNTP concentration was used to examine the only missing site in 18S, U1668. As expected, a stop corresponding to that modification site was not detected (Figure 15). It is interesting to note that this site was also not detected using Ribometh-seq in HeLa cells (Krogh et al. 2016). Further evidence from more cell lines will be required to confirm whether this site is actually modified in other cells or tissues.

Figure 14. Volcano plot of the –log10 p value vs log2 fold change.

Each dot represents a single base position in 18S and 28S rRNAs. Base positions are artificially filtered by p values and log2 fold changes and color coded. Red dots represent positions with log2 fold change <= 7 and adjust p value > 10e-7. Teal dots represent log2 fold change <= 7 and adjust p value < 10e-7. Green dots represent log2 fold change > 7 and adjust p value > 10e-7. Finally, purple dots represent log2 fold change > 7 and adjust p value < 10e-7. positions labeled with purple are determined as high confidence sites. Zoomed in view for two regions are examples to show the actual sites the dots represent. Volcano plot was plotted using R package ggplot2. The dashed red lines indicate manually set cutoffs.

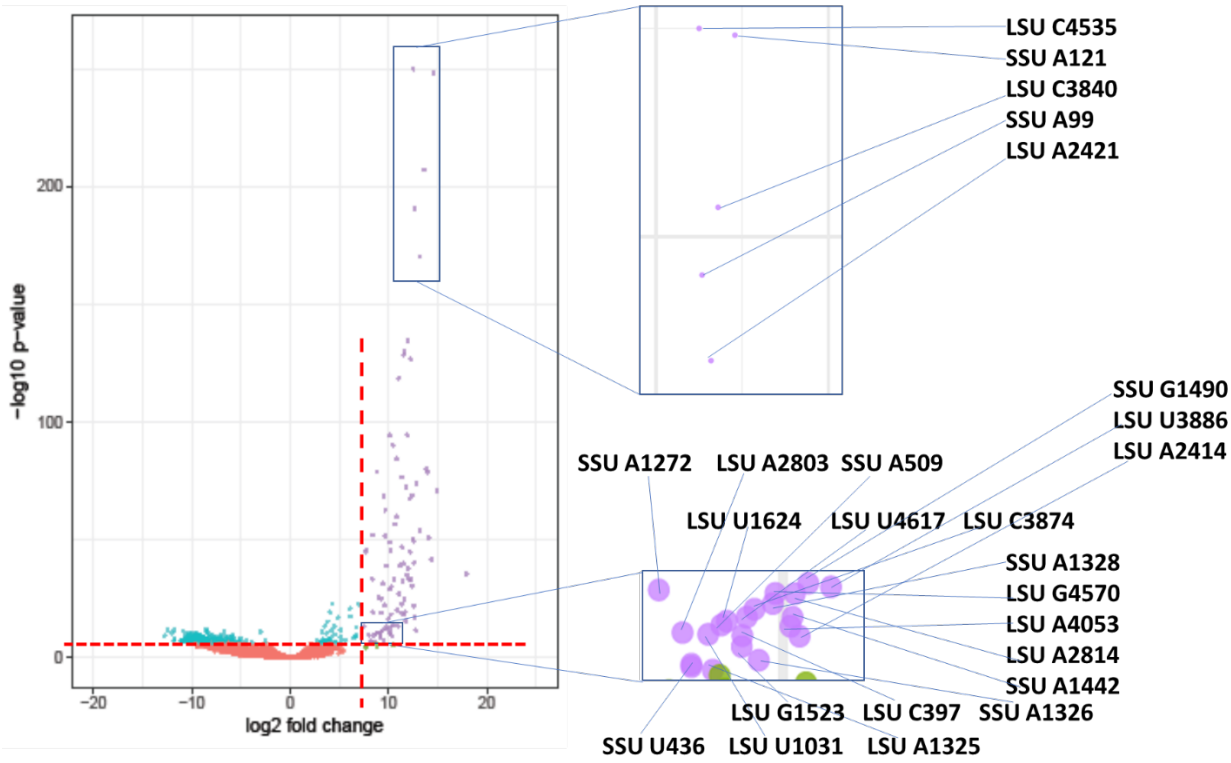
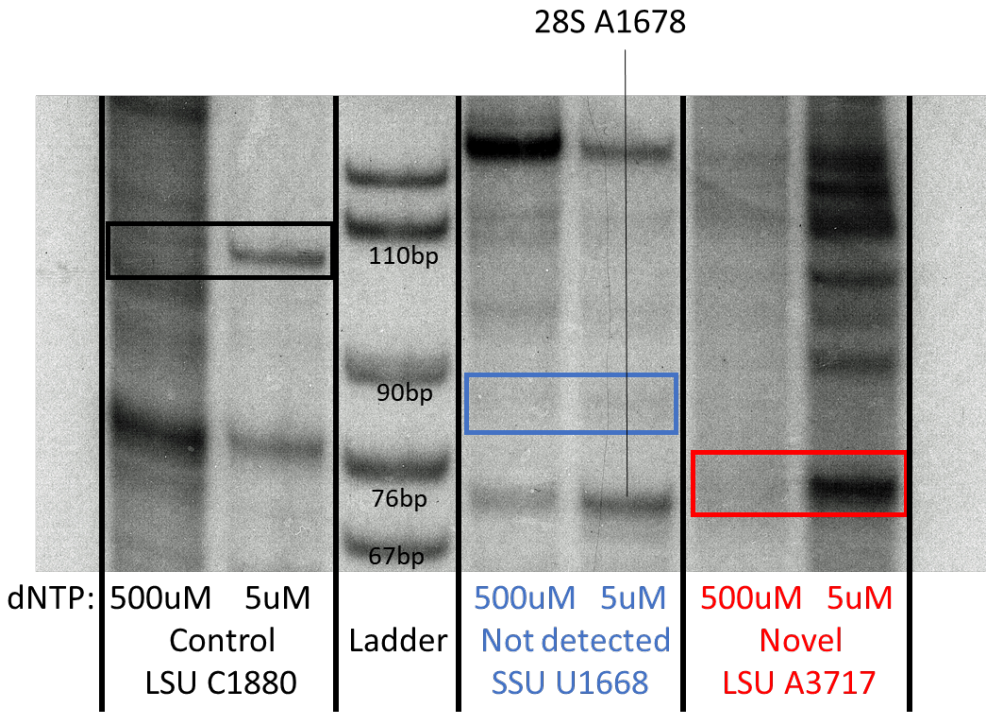

LSU C4535
SSU A121
LSU C3840
SSU A99
LSU A2421
SSU G1490
LSU U3886
LSU A2414
SSU A1272
LSU A2803
SSU A509
LSU U1624
LSU U4617
LSU C3874
SSU A1328
LSU G4570
LSU A4053
LSU A2814
SSU A1442
LSU G1523
LSU C397
SSU A1326
SSU U436
LSU U1031
LSU A1325
−log10 p−value
log2 fold change
200
100
0
−20
−10
0
10
20

Figure 15. Radioactive dNTP concentration dependent primer extension experiment.

The primer extension is performed for three selected sites for RibOxi-seq: positive control at LSU C1880 (lanes 1 and 2), known but not detected SSU U1668 (lanes 4 and 5) and newly detected and not previously known LSU A3717 (lanes 6 and 7). First lane of each set is negative control where primer extension is performed with regular dNTP concentration. Second lane of each set is performed at low dNTP concentration to allow capture of extension stoppage.


28S A1678
110bp
90bp
76bp
67bp
dNTP: 500uM 5uM
Control
LSU C1880
Ladder
500uM 5uM
Not detected
SSU U1668
500uM 5uM
Novel
LSU A3717

3. RibOxi-seq requires modest input material but not high sequencing depth.

Accurate determination of sites of 2'-OMe using RibOxi-seq relies not only on peak calling of oxidized samples, but also on statistical comparison of signals between oxidized and control lanes. An initial pilot experiment using a small quantity of total RNA in combination with Illumina MiSeq sequencing generated ~2.5 million reads for each sample (note: the actual number of aligned reads was much lower). Upon examining alignment with 3'-end only reporting, the pattern was strikingly consistent between experiments, with known sites across 18S and 28S rRNAs represented by strong peaks in oxidized samples with corresponding gaps in control samples mapping to the known sites. After single-base differential expression analysis, 36/40 sites in 18S and 54/66 sites in 28S were detected using a filtering strategy similar to that described above. However, there were also more than 30 new sites detected (Data not shown). Those were likely false positives owing to a lack of enough total available control sample 3'-base counting reads for DESeq2 statistical analysis. Thus, while promising, this pilot experiment was not good enough for accurate peak calling. In our experience, highly sensitive and accurate site detection is achievable at ~12 million reads (sequencer output) per sample.

4. RibOxi-seq is sensitive to methylation changes that are induced by complete depletion of specific snoRNA guides.

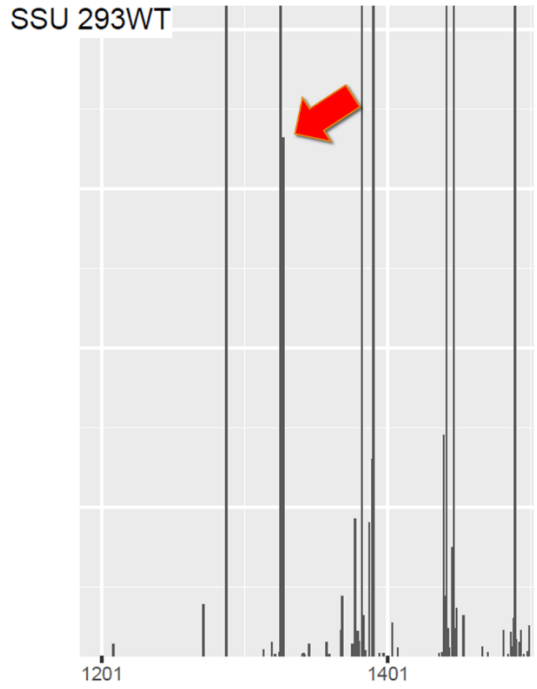
To validate that our method is sensitive to methylation changes, we performed RibOxi-seq on 293T RNA with snoRNA U63 KD. Although U63 directs methylation on 28S rRNA at position 4541 A, we found no major differences in the methylation status between PA1 and 293T cell lines. We speculated that this is because KD experiments do not completely deplete snoRNAs, and it is possible that a small fraction of the total amount is capable of directing normal levels of methylation. We then sequenced additional samples with either U32 snoRNA

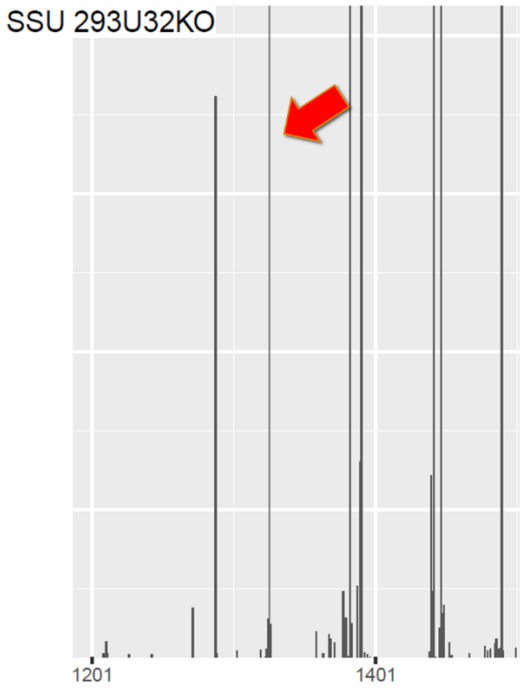
KO or U32+U51 snoRNA DKO. U32 is known to direct methylation at both positions 18S G1328 and 28S A1523, while U51 is known for 28S A1523. Consistently, no 2'-OMe signals were detected at 18S G1328 for U32 KO sample, and 28S A1523 signal is retained. When both snoRNAs were knocked out, neither site was detected by RibOxi-seq (Figure 16). these results confirm the robustness of Riboxi-seq in detecting 2'-OMe sites.

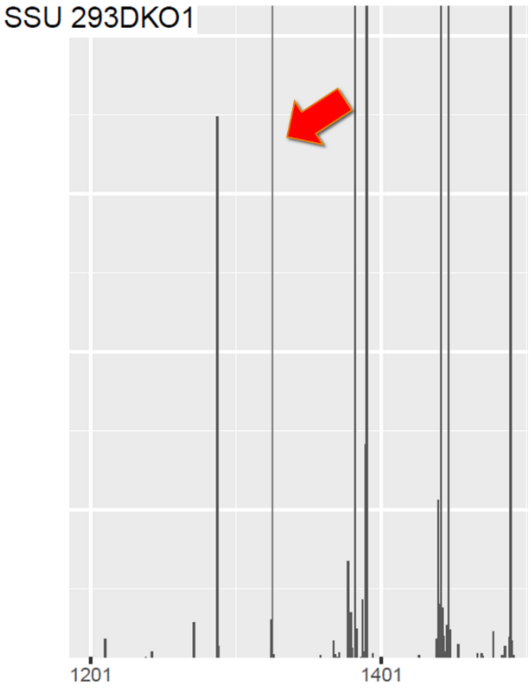
Figure 16. RibOxi-seq on snoRNA KO 293 total RNA samples.

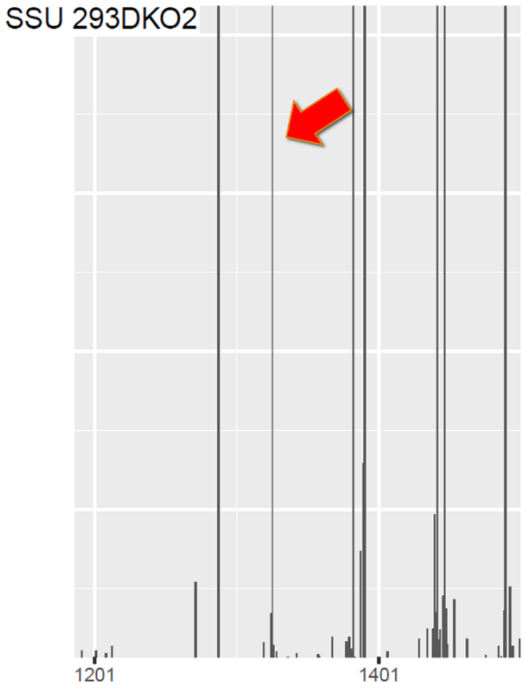
Track 1 vs. track 2 vs. track 3 vs. track 4. On 18S rRNA, track 1 represents WT sample and has a strong 2'-OMe peak at position 1328. U32KO and both double KO sample lost that peak.

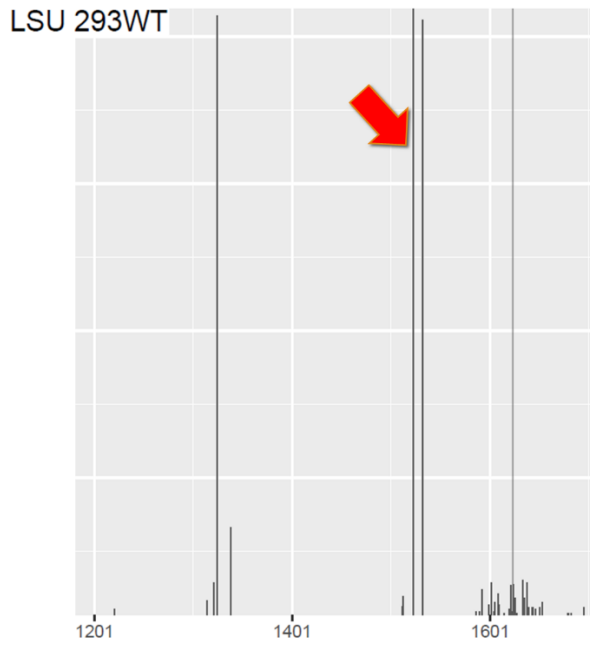
Track 5 vs. track 6 vs. track 7 vs. track 8. On 28S rRNA, track 5 represents WT sample and has a strong 2'-OMe peak at position 1523. U32KO also has the same peak present. Both DKO samples lost the peak.

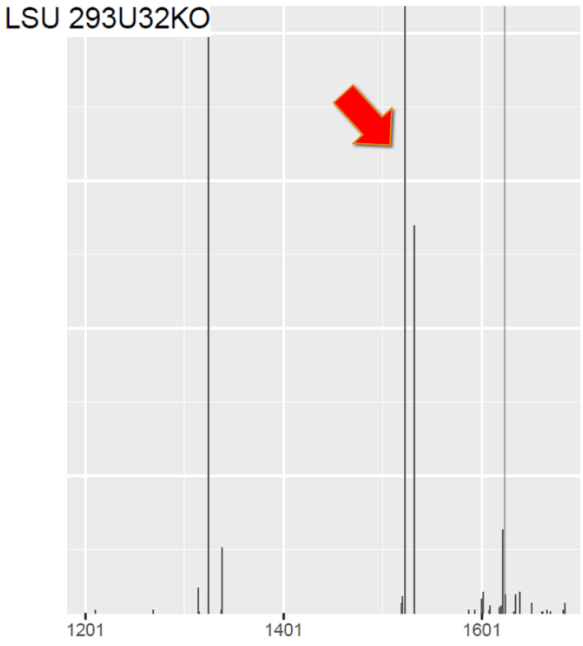

SSU 293WT
1201
1401

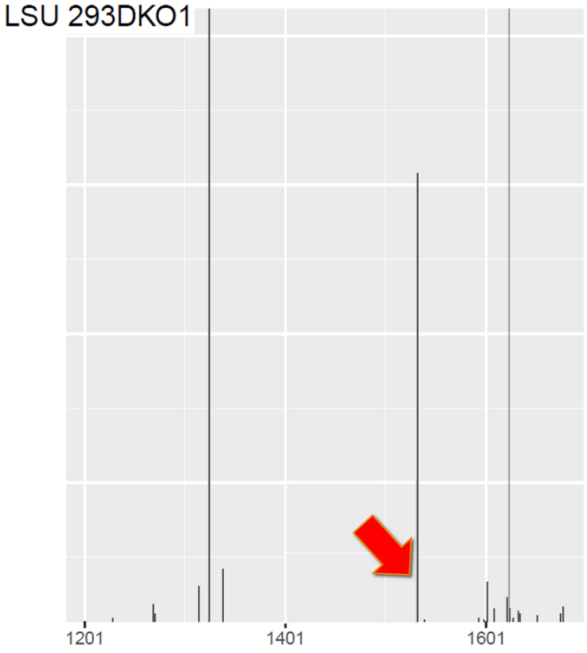

SSU 293U32KO
1201
1401

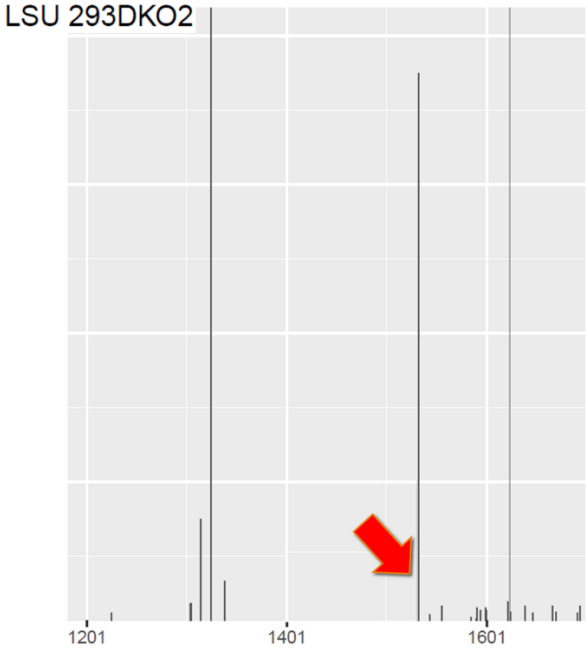

SSU 293DKO1
1201
1401


SSU 293DKO2
1201
1401


LSU 293WT
1201
1401
1601


LSU 293U32KO
1201
1401
1601


LSU 293DKO1
1201
1401
1601


LSU 293DKO2
1201
1401
1601

E. Discussion

Here we have demonstrated that the principle of using Sodium Periodate to enrich 2'-OMe sites are feasible. By coupling the chemistry with Illumina RNA sequencing, we successfully developed a highly sensitive and accurate method to profile 2'-OMe. The wide applicability as evidenced in experiments using multiple cell lines makes it suitable for both explorative and non-quantitative validation purposes. We verified that the method is sensitive to methylation status changes through snoRNA KOs, and even showed RibOxi-seq's capability of detecting methylation variabilities within a cell line.

Ribose methylation occurs on average about 1 in every 60 nucleotides in 28S and 18S rRNA. Under such conditions, one round of β-elimination is sufficient for accurate site detection even at low sequencing depth. However, the occurrence of methylation is very likely to be far lower in other RNAs such as lncRNAs and mRNAs, and no instances of this modification have been reported so far in human, other than in 5'-cap structures. In order to detect mRNA modifications, not only will higher sequencing coverage be required, but also perhaps multiple β-elimination steps to greatly increase the probability of 3'-end 2'-OMe exposure. Also, additional β-elimination can be used on rRNAs if the amount of input total RNA (~7.5 μg per sample) described in the standard protocol is impossible to obtain. The required starting material can be divided by 2–4 for each round of β-elimination added.

Several limitations still exist. Although the number of sequencing reads required is significantly lower than that for Ribometh-seq, the input RNA material required is in general somewhat higher, at the micrograms level, with the possibility of reduction to a sub-microgram level if using additional β-elimination steps. Also, as described above, owing to the nature of ribose methylated bases being resistant to nuclease and alkaline hydrolysis, it is difficult to detect

adjacent modified bases if the distal base is fully methylated. Further, exposure of methylated bases relies on extensive and random digestion. tRNAs and other RNAs shorter than 100 bp are difficult to study because the sizes of fragments that would need to be generated might be quite small and difficult to examine. Finally, the method at the current stage cannot be used as a quantitative tool to compare methylation intensity between different sites, since linker ligation efficiencies using T4 RNA ligases 1 and 2 have been described to have sequence biases (Raabe et al. 2014). Hence, without first determining ligation efficiencies of linkers to each of A, U, C, and G bases using spike-ins of known 2'-O-methylated oligos as internal controls, comparisons between different sites is not yet possible. On the other hand, comparison between different samples at the same site appears to be feasible.

F. Complete protocol

The library preparation using our standard protocol requires around 4–6 days to complete with a moderate work load, factoring out optimization steps. However, the procedure can be stopped whenever an ethanol precipitation is performed and the sample is resuspended into nuclease-free water. Alternatively, samples can also be left precipitating in 100% EtOH under −20°C indefinitely to increase yield.

RNA extraction

For cells cultured in 10 cm Petri dishes, a PureLink RNA Mini Kit is used in conjunction with the PureLink on-column DNase set to isolate RNA and remove genomic DNA. Steps for the extraction are detailed in the Purelink Kit's protocol. In case any overexpression system is used, additional post-extraction gDNA removal may be necessary due to an increase in DNA molarity.

The TURBO DNA-free Kit has proven effective for such conditions after following its "Rigorous DNase treatment" protocol.

1. Extract total RNA using the PureLink RNA Mini Kit with PureLink on-column DNase.

2. Optional: Use the Turbo DNase Kit to further remove contaminating DNA.

RNA fragmentation

Each sample set should have at least one control and one oxidation sample. We recommend a minimum of three technical replicates for each. After completing fragmentation and the subsequent two steps until just before the oxidation procedure, the RNA loss should be ~40%. It is recommended to use an initial total RNA amount of 30 μg (can be lowered upon further optimization), which will yield about 18 μg of fragmented RNA (results may vary for each laboratory). The amount of fragmented RNA recommended for oxidation is 4–6 μg, and 1 μg for nonoxidation. Three technical replicates for all samples require about 15 μg total. The amount of starting RNA can be significantly lowered if using the alternative procedure during oxidation and β-elimination steps. Such an alternative is also necessary if applying the method on other samples depleted of rRNAs.

3. In a 1.5 mL Eppendorf tube, dilute 15 μg of total RNA (for each sample) into 133.5 μL with nuclease-free H_2O.

4. Vortex the mixture and spin for a brief second to collect all liquid at the bottom.

5. Place the tube into a 90°C heat block for 3 min to denature the RNA and immediately place on ice for at least 1 min.

6. Dilute 1 μL stock Benzonase (341 U/μL) into 500 U/mL using 681 μL of 1× Benzonase buffer. (Always dilute fresh prior to using. Do not freeze.)

7. Add 15 μL of 10× Benzonase buffer and 1.5 μL of diluted Benzonase to the diluted RNA (final RNA concentration: 100 ng/μL). Incubate on ice for 90 min. (This incubation time only serves as a starting reference since it can greatly vary between different laboratories. Perform a time-course experiment to establish optimal incubation time based on desired fragment length.)

8. Perform phenol–chloroform extraction. Add 150 μL acid phenol:chloroform and vortex for 10 sec. Centrifuge at 20,000g at RT for 5 min.

9. Transfer about ~150 μL of the supernatant into a new set of 1.5 mL Eppendorf tubes.

10. Ethanol precipitate the RNA. Add 30 μL of 3 M sodium acetate (0.1× volume), 120 μL water. Mix well and add 750 μL of 100% ethanol (2.5× volume).

11. Mix well and place on ice for >30 min to precipitate RNA.

12. Spin at max (>16,000g) at 4°C or RT for ~30 min.

13. Carefully remove the ethanol without dislodging the pellet.

14. Add ≥500 μL 70% ethanol, vortex, and centrifuge for 10 min to wash the pellet.

15. Carefully remove the ethanol and air dry the pellet for 2 min.

16. Resuspend pellet in 100 μL of nuclease-free H_2O.

17. Optional: Examine fragmented RNA size distribution using TapeStation 2200 and RNA Screen Tapes. Ensure the pattern has a strong peak spanning the region from 50 to 200 bp. For

poly A + samples, it is recommended to increase digestion strength to shift the peak toward even smaller sizes.

18. Ethanol precipitate 5–6 µg fragmented RNA of each sample replicate.

Oxidation

To prepare the fragmented RNAs for subsequent elimination of the 3'-end bases to expose ribose methylated bases that are potentially positioned one base upstream, these ends need to be oxidized into dialdehydes using $NaIO_4$.

19. Freshly prepare 200 mM $NaIO_4$ solution by dissolving 42.78 mg of the $NaIO_4$ powder per 1 mL oxidation buffer. Protect the solution from light. Important: This step should be performed while precipitating the fragmented RNA and not earlier.

20. Dissolve RNA pellet in 30 µL oxidation buffer (make sure the pellet is well resuspended) and add 10 µL of prepared $NaIO_4$ solution. Mix well and incubate at room temperature protected from light for 45 min (briefly vortex and spin the reaction tube every 15 min to ensure proper resuspension of the pellet).

21. Adjust the volume to 90 µL with nuclease-free H_2O. Add 10 µL H_2O and 1 µL LPA.

22. Perform ethanol precipitation. Important: After adding 100% ethanol for precipitation and centrifugation, white precipitates/film might be seen scattered on the Eppendorf tube wall due to high salt content and a typical pellet might not be visible. This is normal. RNA can be recovered if the orientation of the tube is kept consistent and a pipettor is used to vigorously flush and/or gently scrap the precipitates/film off the wall after adding elimination buffer.

β-Elimination

This step catalyzes the leaving of the 3'-end oxidized base to expose potentially 2'-O-methylated bases at the 3'-end of the fragments.

23. Add 50 µL of β-elimination buffer to dissolve the pellet. Vigorously vortex, pipette up and down, or invert the tube to further resuspend the oxidized RNA.

24. Spin the tubes briefly and transfer samples into PCR strip tubes. Tubes with individual caps are highly recommended to prevent cross-contamination, especially in later steps where control samples are handled alongside oxidation samples.

25. Using a thermal cycler, incubate at 45°C for 85 min.

26. Process samples through NucAway spin columns to remove unwanted small fragments and salts from the samples (β-elimination alkaline conditions which can generate small undesirable fragments).

27. Transfer samples into a new set of 1.5 mL Eppendorf tubes.

28. Ethanol precipitate and resuspend in 22 µL H_2O as in the previous oxidation step. Adding additional 0.5 µL LPA is highly recommended.

Phosphate removal and oxidation

To oxidize all 3'-ends that are not 2'-O-methylated, another $NaIO_4$ treatment is needed. However, after β-elimination, RNA fragment 3'-ends will contain a mixture of 3'-phosphates, 2'-phosphates, and/or 2'-3'-phosphates. It is vital to remove these phosphate groups to avoid false positives in the final data representation. T4 PNK is used to remove all three types of phosphates.

29. Transfer to PCR tubes.

30. Remove 3'-, 2'-, and 2'–3'-cyclic phosphate using T4 PNK. (Important: The PNK buffer used must not contain ATP.)

Components	Volume (μL)
β-Elimination treated RNA	22
2× PNK buffer diluted from the included 10× buffer (pH adjusted to 6.0)	25
SUPERase• In	1
NEB T4 PNK	2

31. Incubate at 37°C for at least 4 h (longer incubation times may be OK and might be beneficial as T4 PNK is an inefficient phosphatase).

32. Important: here is where control and oxidized samples are separated. Phenol–chloroform extract the dephosphorylated RNA and aliquot 20% of the supernatant for each sample as control, while the remaining 80% will go through final oxidation.

33. Ethanol precipitate controls and resuspend them with 16 μL H_2O; at the same time, ethanol precipitate oxidation samples into pellets (add 1 μL LPA for both). Store the controls in -80°C.

34. Freshly prepare 200 mM $NaIO_4$ solution by dissolving 42.78 mg of the $NaIO_4$ powder in 1 mL oxidation buffer. Protect the solution from light.

35. Dissolve the RNA pellet from oxidation samples thoroughly in 30 μL oxidation buffer.

36. Add 10 μL of prepared $NaIO_4$ solution. Mix well and incubate at room temperature protected from light for 45 min.

37. Transfer to a new 1.5 mL Eppendorf tube and ethanol precipitate RNA and resuspend in 16 μL H_2O. If white precipitates remain in the solution, do not remove them.

38. Following these steps, the samples are properly oxidized.

3’ DNA linker ligation

Ligation of 3’ linkers to unoxidized RNA 3’-ends will enable selective reverse transcription and thus, enrichment of 2’-O-methylated RNA fragments. From this step on, control samples will be subjected to the exact same procedures.

39. Thaw the control samples. Transfer 8 µL for each sample into PCR tubes.

40. Transfer 8 µL of each 16 µL oxidized sample into PCR tubes. Store the remaining 8 µL of both control and oxidation for each sample at −80°C as backup.

41. Ligation of 3’ linker (prepare the reagents in a way that can be properly mixed, as the amount of PEG 8000 added can make it difficult).

Components	Volume (µL)
Oxidized/control RNA	8
3′ DNA linker 50 µM (final 2.5 µM)	1
RNaseOUT	1
50% PEG8000	7
10× t4 RNA ligase buffer	2
T4 RNA ligase 2 truncated kq	1

42. Incubate the reaction in thermal cycler at 16°C overnight for 18 h.

Anneal RT primer

Any 3’ linker not ligated in the previous step will still be freely available for ligation in the samples. To ensure the 5’ RNA linker ligation attaches the RNA linkers to sample RNA fragments but not to the free 3’ linkers during the next step, the RT primer is annealed first.

43. Add 1 µL of the 50 µM RT primer and 69 µL nuclease-free water to each sample.

44. Incubate in thermal cycler with the following program:

90°C for 2 min

65°C for 10 min

4°C for 1 min

45. Phenol–chloroform extract and ethanol precipitate (add 1 µL of LPA) each sample and resuspend in 11 µL H_2O.

5' RNA linker ligation

The double-stranded structures resulting from annealing the RT primers and free 3' linkers will prevent them from being ligated to the 5' RNA linkers (Munafó and Robb 2010).

46. Thaw 50 µM RNA linker from −80°C and transfer (number of samples) *1.3 µL into a PCR tube.

47. Denature RNA linker at 72°C for 2 min and return to ice.

48. Prepare the following ligation reaction.

Components	Volume (µL)
Annealed RNA	11
5′RNA linker 50 µM (final 2.5 µM)	1
100% DMSO	2
10× t4 RNA ligase buffer	2
10 mM ATP	2
RNaseOUT	1
T4 RNA ligase 1	1

49. Incubate at 25°C for 1 h, then terminate reaction at 65°C for 15 min.

Reverse transcription

Because the RT primer has been annealed in the previous step, the RT reaction can proceed without addition of primer. During this step, cDNA is synthesized. At the same time, random-hexamer sequences built into the RT primer are also incorporated into the cDNA library. These will allow the removal of PCR duplicates later during data treatment.

50. Prepare RT reactions to generate a cDNA library using the following setup with SuperScript III included reagents.

Components	Volume (µL)
Ligated RNA	19
SuperScript III RT buffer	4
$MgCl_2$	8
DTT	4
dNTP mix	2
RNaseOUT	1
SS III enzyme	2

51. Incubate the reactions in thermal cycler following the kit's protocol.

52. Hydrolyze remaining RNAs by adding 4.4 µL of 1 N NaOH and incubate at 98°C for 20 min.

53. Add 22 µL 200 mM Tris-HCl PH = 7.0 to neutralize the PH.

54. Use Ampure XP beads at 0.8:1 ratio (add 53 µL Ampure XP solution). Incubate for 5 min to let beads bind cDNA of sizes 250 bp and above.

55. Transfer supernatant to new tubes (discard beads) and add an additional 67 µL Ampure XP solution to make the ratio 1:1.8. Finish Ampure XP purification.

56. Elute using EB buffer.

PCR amplification of cDNA library

Illumina i5 and i7 PCR primer sequences have sequences complementary to the 5' RNA linker and RT primer sequences. This allows direct PCR amplification of the cDNA library. Periodate oxidation enrichment in previous steps results in a cDNA library of very low complexity. Hence it necessitates additional amplification cycles compared to construction of other types of sequencing libraries (~35 cycles versus ~12 cycles). The strandedness of the final library is second-strand similar to the library prepared using the Illumina ligation method.

57. Prepare NEB Q5 PCR reactions.

Components	Volume (µL)
cDNA	8.5
I5 primer	2
I7 primer	2
Q5 2× master mix	12.5

58. Incubate in the thermal cycler using the following program modified from the Q5 protocol. e library prepared using the Illumina ligation method.

Step	Temp	Time
Initial denaturation	98°C	30 sec
One cycle	98°C	5–10 sec
	55°C	10–30 sec
	72°C	20–30 sec/kb
32 cycles	98°C	5–10 sec
	62°C	10–30 sec
	72°C	20–30 sec/kb
Final extension	72°C	2 min
Hold	4–10°C	

59. Add 25 µL of AmpureXP to achieve a 1:1 ratio to select for fragment of sizes ~200 bp and above, reducing the amount of non-insert fragments (Illumina i5 sequence: ~70 bp, Illumina i7 sequence: ~66, total non-insert product: ~136 bp).

60. Purify libraries following the AmpureXP protocol. Resuspend each sample in 15 µL with Illumina RSB (resuspension buffer).

Library quantification

61. Use TapeStation and DNA Screen Tape to visualize library size distributions. It should have a distribution around 200 bp. Even though stringent steps have been taken to avoid non-insert PCR products, they may still be present and are expected.

62. Accurately measure library concentrations using the Qbit Fluorometer and dsDNA High Sensitivity Reagent Kit (follow Qbit protocol).

63. Calculate library molarity using sizes and concentrations measured.

Sequencing

The NextSeq 500 (or MiSeq) 150 cycle Mid Output Kit is used in this experiment with 75 bp by 75 bp configuration.

64. Prepare and load the libraries for the NextSeq 500 sequencer following the established protocol. Important: Make sure PhiX phage DNA comprises at least 30% of the total library if the Riboxi-seq samples are the only samples being sequenced because of the nature of extremely low diversity amplicon sequencing. The final loading concentration used in this experiment is 1.5 pM, which is slightly lower than the 1.8 pM from the protocol.

Data treatment

65. Remove read-through sequences. Because of the nature of the sequencing library preparation, many fragments will have insert sizes significantly smaller than the 75 bp length sequenced by the sequencer. The resulting read-through sequences will greatly impact alignment reliability. Cutadapt (Martin 2011), a Python package, is used to first remove read-through sequences at the 3'-end for both read 1 and read 2. An in-house script has been used to compare randomer

sequences and ~5 bp RNA sequences to determine and collapse PCR duplicates on read 2. Finally, we used cutadapt to remove the randomers and linker sequence from 5'-ends of read 2. The resulting ".fastq" files can then be used for alignment.

66. TopHat2 is used to align the reads. The annotation file used consists of chrUn coordinates, which contain two sets of 18S and 28S rRNAs, extracted from the hg19 index (Kim et al. 2013).

67. The "accepted_hits.bam" for each sample are sorted using SAMtools and converted to ".BED" files using BEDtools (Heng et al. 2009; Quinlan and Hall 2010).

68. The third column of the ".BED" file represents the starting and ending positions (with respect to 5'- and 3'-ends) of each read, while the sixth column indicates sense or anti-sense. We counted the number of reads with 3'-end alignment for each position corresponding to 18S and 28S rRNA and generated a count table for each sample (only sense reads are taken, since the two rRNAs we aligned to the reference sequence were transcribed from the sense strand).

69. The count data files were then imported into DESeq2 for differential analysis per base position (Love et al. 2014).

All raw and processed data produced from studies in this chapter II were deposited at the NCBI Gene Expression Omnibus (GEO) with accession number: GSE96999.

All raw and processed data pertaining to *T. brucei* RibOxi-seq produced from experiments in chapter III were deposited at the NCBI Gene Expression Omnibus (GEO) with accession number: GSE102516 pending availability of corresponding publication.

Chapter III

Application and Extension of RibOxi-seq

A. Abstract

The successful development of the RibOxi-seq provides us the foundation of an accurate, robust and cost-effective way to studying 2'-OMe on rRNAs. Although from an evolutionary perspective, *T. brucei* is very distant from human within the Eukarya domain, the conservation of 2'-OMe machineries is an interesting observation. Although there have been a large number of snoRNAs identified, as well as bioinformatically predicted corresponding 2'-OMe sites on *T. brucei's* rRNAs, the number of experimentally validated sites remains much smaller. Besides the field of epitranscriptomics gaining attraction, more evidence suggests an increasing importance of 2'-OMe in different life stages of *T. brucei*. Therefore, we sought to profile 2'-OMe landscapes of different stages of the organism to help better understanding the biology of its lifecycle and to further highlight the power and potential of RibOxi-seq. In addition, we would like to be able to answer the questions asked in Chapter I such as: are there internal 2'-OMe sites within mRNAs and can we identify targets for orphan snoRNAs? We performed pilot experiment on mouse mRNA, as well as human ESC WT mRNA and SNORD116 KO mRNA. In both human and mouse, RibOxi-seq suggests a strong likelihood of detecting 2'-OMe sites, pending validation with an independent method, and shows interesting patterning of the modification distribution. With some technical difficulties still present, we are still actively exploring ways to analyze and compare WT and KO data.

B. Background

Trypanosomes. *T. brucei* is a protistan parasite causing deadly disease in sub-Saharan Africa. During its life cycle the parasite shuttles between its tsetse fly vector and human host where it lives freely in the bloodstream. For a rather small genome (35Mb haploid vs. human: 3200Mb diploid), trypanosomes encode a large repertoire of snoRNAs. (Schultz et al. 2006, Uliel et al. 2004). Although there have been continuous efforts made to map sites of 2'-OMe in trypanosomes, only a few laboratories around the world have pursued RNA modifications study in parasitic organisms. Over the years, evidence for individual sites slowly accumulated mainly contributed by targeted primer extension studies; however, a lack of curated reports or databases documenting known sites that are experimentally validated renders 2'-OMe patterns in *T. brucei* and related organisms elusive. Currently, more than 20 snoRNA clusters encoding more than 50 Box C/D snoRNAs have been identified in transcriptomic studies, suggesting a capability of modifying more than 130 sites in rRNA alone. A significant portion of these 130 sites, unlike human counterparts, are bioinformatically predicted rather than exhaustively validated (Liang et al. 2005, Uliel et al. 2004, Barth et al. 2008).

Interestingly, a recent semi-large-scale primer extension study revealed the possibility of differential 2'-OMe patterns between the mammalian-infective bloodstream and insect-stage procyclic forms (BF and PF) of *T. brucei* (reviewed in Ponte-Sucre 2016). The detection of methylation level changes in certain sites, rather than a change in overall methylation levels, provided evidence for possible functional involvement of 2'-OMe in the life cycle of the parasite (Barth et al. 2008). With the opportunity to form a close collaboration with Dr. Arthur Günzl's lab within our department as well as the Günzl lab's collaborators who are experts in Trypanosome 2'-OMe, we decided that it would be appropriate and feasible to determine and compare rRNA 2'-OMe profiles between *T. brucei* life cycle stages.

Mammalian sites of 2'-O methylation. In Chapter II, I discussed that if we want to adapt RibOxi-seq for detection of low abundance transcripts, such as mRNAs, additional rounds of β-elimination would likely be necessary. However, each round involves several hours of work. We alternatively speculated that if we simply increased digestion time length, the odds of sites on mRNAs being exposed should increase. Even though there will be significant potential loss on number of sites that could have been detected with extra β-elimination, this pilot study only serves as proof-of-principle that there are mRNA internal 2'-OMe and that they can be detected by RibOxi-seq.

Introns of ribosomal protein L13a (Rpl13a) codes for a group of snoRNAs including U32a, U33, U34 and U35a, whose rRNA targets have been known. Interestingly, snoRNAs from the Rpl13a locus are critical regulators of reactive oxygen species and oxidative stress (Michel et al. 2011, Lee et al. 2016). In an animal model of diabetes from drug-induced beta cell injury, genetic loss of the Rpl13a snoRNAs (-/-) reduces oxidative stress, confers partial resistance to development of diabetes, and also leads to recovery from injury beta-cell injury with normalization of blood glucose over time. Although box C/D snoRNAs were not previously known to target mRNA for modification, the Holley Lab has shown preliminary data that these snoRNAs might target mRNA for 2'-OMe modification as a novel mechanism of action (Holley et al. 2015, Holley lab unpublished data). Therefore, we formed a collaboration in the hope of finding direct evidence for the methylation on mRNA and gaining insights on clinical relevance of 2'-OMe.

SNORD116 is a box C/D snoRNA cluster residing within chromosome 15 q11-q13 region. There are 30 similar snoRNAs encoded by the cluster (Cavaillé et al. 2000). Prader-Willi syndrome (PWS) is a genetic disorder resulted from certain deletions within the region. PWS

disrupts normal functions of hypothalamus neurons, including abnormal hormone levels (reviewed in Cassidy et al. 2012). However, the underlying molecular basis of the disease is not well understood. The locus contains several genes, although interestingly, SNORD116 deletion always accompanies PWS patients (Bieth et al. 2015). Thus, the 2'-OMe aspect of the disease have not been explored as the SNORD116 cluster snoRNAs are orphan snoRNAs that lack known targets. Thus, we would like to investigate whether the SNORD116 snoRNAs have 2'-O-methylation targets and whether the methylation process is involved in the pathology by investigating human 2'-OMe patterns transcriptome-wide.

C. Methods and Materials

1. RNA extraction and PolyA RNA isolation

Total RNA from axenic *T. brucei* cultures (both BF and PF were cultured in Günzl lab) were prepared using PureLink® RNA Mini Kit following manufacturer's protocol (with on-column DNAse treatment). Total RNAs from U32KO/WT mice livers were supplied by Holley lab from Duke University. SNORD116 KO/WT H9 human embryonic cell lines were provided by Michael Chung from the Chamberlain lab, and RNA extractions were performed using PureLink® RNA Mini Kit. mRNAs were enriched for mouse and H9 samples using PolyATtract® mRNA Isolation Systems from Promega.

2. RibOxi-seq

Standard RibOxi-seq was performed on both *T. brucei* PF vs. BF total RNA samples, and H9 SNORD116 WT vs. KO total RNA samples with 2 biological replicates each. RibOxiseq-seq protocol was slightly modified in order to improve possibilities for mRNA-site detection. The

modified protocol was performed on the following sample sets with 2 biological replicates each: mouse U32 WT vs. KO mRNAs; and human H9 SNORD116 WT vs. KO mRNAs.

D. Results

1. Profiling of *T. brucei* 2'-OMe rRNA landscape supports life stage-specific and site-specific differential methylation

As mentioned earlier, there has been no curation of experimentally validated *T. brucei* 2'-OMe sites for rRNAs to date. However, snoRNA repertoire has been extensively examined, which also generated a number of 2'-OMe predictions (Michaeli et al. 2012). Currently, Dr. Shulamit Michaeli's lab (one of our collaborators), is consolidating evidence of *Trypanosome* rRNA 2'-OMe from various sources including bioinformatic predictions, targeted primer extension experiments, structural studies, 2'-OMe-seq and known snoRNAs. The goal is to cross check between different 2'-OMe detection methods and corresponding snoRNAs. Eventually, each position will be evaluated based on cumulative evidence and a cutoff will be set for determination of the methylation status. Since RibOxi-seq performed very well on human rRNA site-detection, we were optimistic about its potential contribution to this collaborative effort.

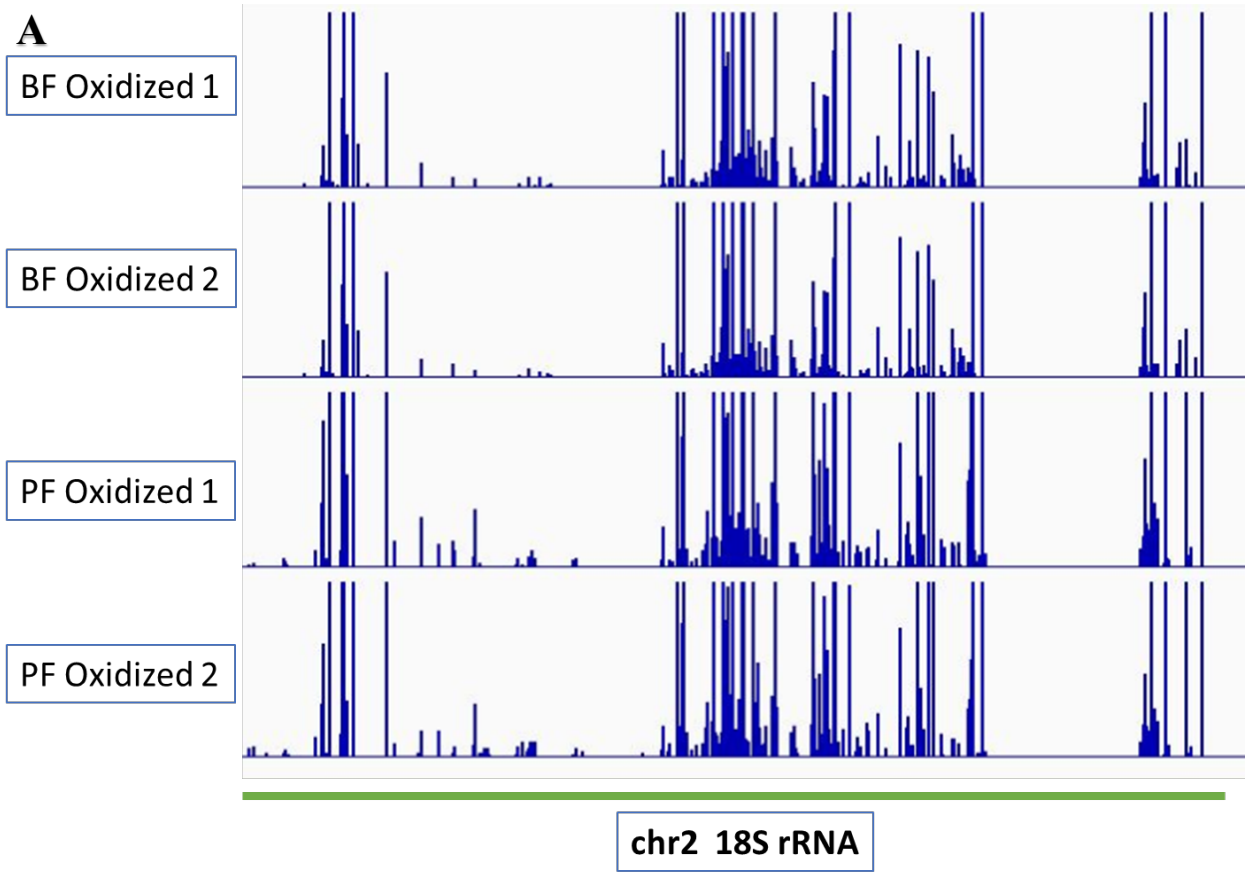
Overall 2'-OMe patterns generated through RibOxi-seq between the two life stages are remarkably similar when visualized on the IGV genome browser (Figure 17A). Interestingly, after filtering the DESeq2 statistical analysis output list with adjusted p-value<0.000001 and log2 fold change>6, we observed highly comparable high-confidence sites between the two life stages with minor differential methylation (Figure 17B). When comparing to data in the lists currently being compiled as well as the data from MIchaeli et al. (2012), our results experimentally confirmed many of the predicted sites, while validated known sites in a single experiment (Table 8). One interesting observation we made comparing to the compiled data was

that many sites, which are both validated and detected by RibOxi-seq, showed no evidence in existing bioinformatic predictions, underscoring the limitations of algorithms of current tools. When compared to 2'-OMe-seq data, there are a number of agreements and disagreements. However, we argue that our method should represent a rather accurate picture of the 2'-OMe sites because the principle of the chemistry of our approach. Regardless, there are a few candidates that received strong scoring for being potential sites and which are differentially methylated between life stage differences and worth following up using independent approaches (Table 5). How relevant and important differential methylation is on these sites is for *T. brucei* biology will require further investigation.

Figure 17. RibOxi-seq on *T. brucei* BF and PF total RNA samples.

(A) Visualization of read alignment of RibOxi-seq data for both BF and PF. Region selected is 18S rRNA region. Genome build is Tb427. Each track represents 2'-OMe patterns of the two replicates of the BF and PF respectively.

(B) After statistical analysis, example of sites that have significant differential methylation observations. Purple peaks represent positions that have $\log_2$ fold change of greater than or equal to 6, while adjusted p value is less than $10e^{-7}$; Blue peaks represent positions that have $\log_2$ fold change of less than 6, while adjusted p value is less than $10e^{-7}$; And Blue peaks represent positions that have $\log_2$ fold change of greater than 6, while adjusted p value is greater than $10e^{-7}$. In general, only the purple positions are considered once cutoffs are set. Here we selected a few sites that are differentially detected between the two life cycle forms.


A
BF Oxidized 1
BF Oxidized 2
PF Oxidized 1
PF Oxidized 2
chr2 18S rRNA

B

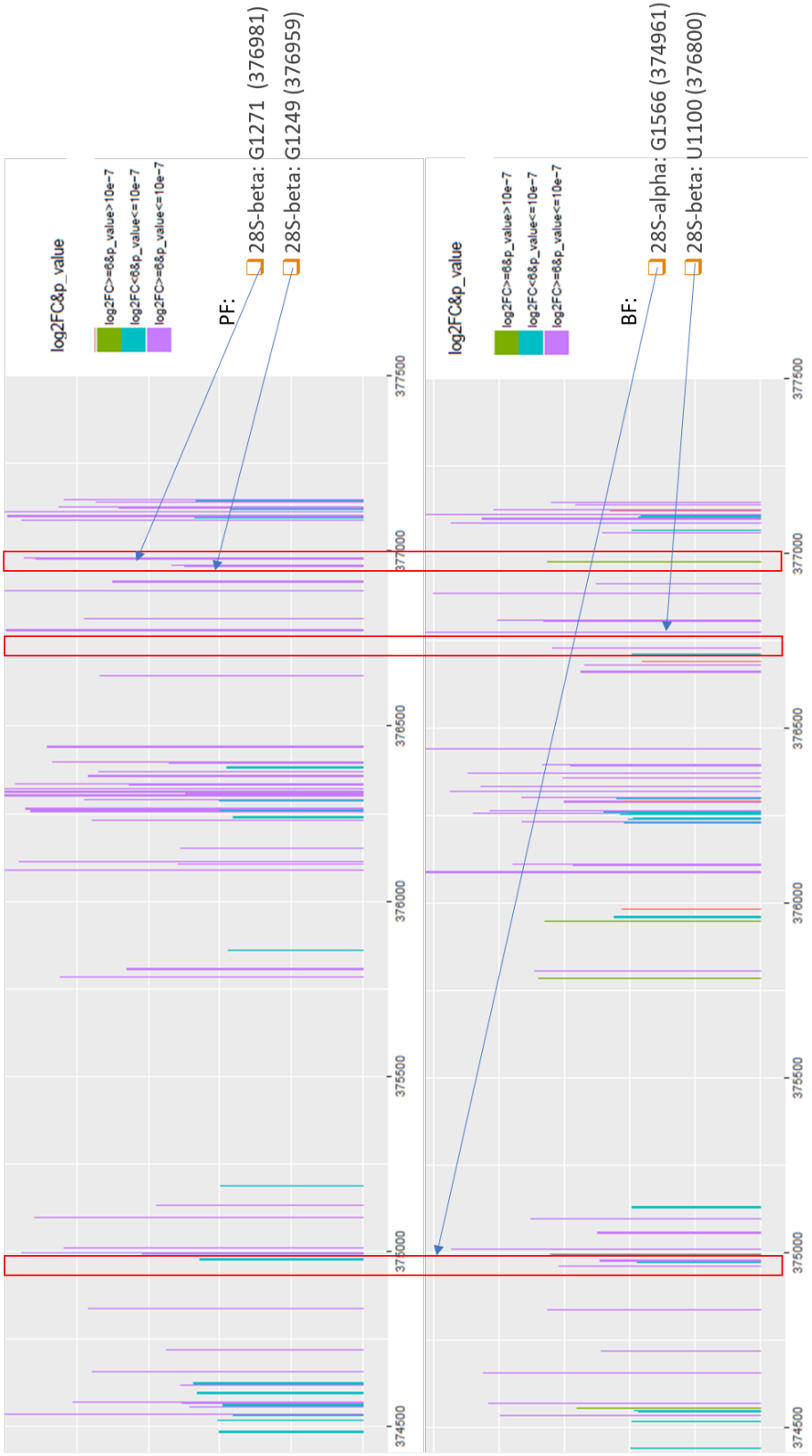

log2FC&p_value
log2FC>=6&p_value>10e-7
log2FC<6&p_value<=10e-7
log2FC>=6&p_value<=10e-7
PF:
28S-beta: G1271 (376981)
28S-beta: G1249 (376959)
BF:
28S-alpha: G1566 (374961)
28S-beta: U1100 (376800)
377500
377000
376500
376000
375500
375000
374500

Table 5. List of promising differentially methylated sites in *T. brucei* to follow up.

rRNA	Base position	Modification	Box C/D snoRNA	Prediction	Primer extension	2'-OMe-seq	RibOxiseq
LSU5	1742	Um	TB7Cs2C1	Yes	?	Yes	Yes, PF Only
LSU5	916	Um	TB11Cs4C2	Yes	Y	Yes	Yes, PF Only
LSU3	601	Cm	TB10Cs1C1	Yes	Y	Yes	Yes, PF Only
LSU3	1264	Cm	TB6Cs1C1	Yes	?	Yes	Yes, PF Only

These sites were confirmed in bioinformatic prediction based on snoRNA sequence, validated using primer extension experiment, detected using 2'-OMe-seq, and detected in RibOxi-seq. Specifically, RibOxi-seq only detected them in PF samples.

2. RibOxi-seq can detect mRNA 2'-OMe sites for genes abundantly expressed in corresponding cell lines or tissues.

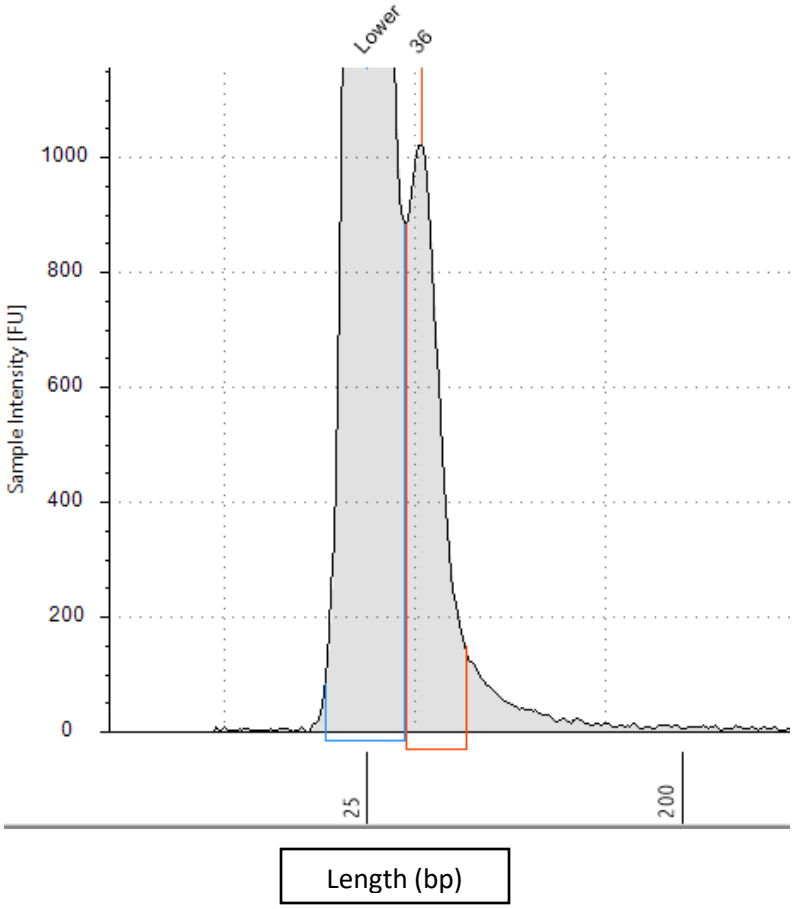
Our original speculation was multiple extra β-elimination steps would be required. However, we were concerned that if starting material is limited, each additional step, which needs to be carried out under alkaline conditions and at elevated temperature, could result in accumulation of RNA breakages and thus extensive loss of material. We expect occurrences of 2'-OMe sites on mRNAs are much less frequent events than on the abundant and highly structured rRNA, which could lead to very little material for reverse transcription after oxidation of RibOxi-seq procedures. With a low RNA abundance RT reaction, it has been known that strong background noise and artefacts can occur (Levesque-Sergerie et al. 2007). Most specifically mis-priming, which is resulted from primer partial hybridization with non-specific targets, can generate 'fake' peaks. In this case, all sequences upstream of detected peaks would appear to contain a motif having high to perfect resemblance to part of the RT primer (Gillen et al. 2016). Such patterns were evidenced in a methodology, using same principles, that was published the same time as RibOxi-seq. Significant mis-priming events consisted of almost 40% of their mRNA sites obtained from initial RNA fragments that underwent 13 additional rounds of elimination (Dai et al. 2017). Also, total RNA, rather than poly(A)$^+$ RNA, was used in the protocol, leading to the possibility of further complications. Interestingly, the *corrigendum* published later in an attempt to address the mis-priming artefact generated data-sets that have no overlap with the original published data-sets, even the 60% of sites that reportedly showed little similarity to the sequencing primer (https://www.nature.com/articles/nmeth0318-226c).

To circumvent this potential pitfall, we argue that if we purify poly(A)$^+$ RNAs and perform very extensive digestion to increase 2'-OMe exposure after a single step of beta elimination, we

may be able to detect at least some sites in mRNAs. Thus, from purified mouse liver poly(A)$^+$ RNAs and human H9 poly(A)$^+$ RNAs, we modified the digestion temperature to generate RNA fragment pools with an average length of 36bp, while the remaining steps of the RibOxi-seq were unchanged (Figure 18). Because rRNAs are highly abundant (making up more than 95% of total RNAs), even with poly(A)$^+$ RNA purification, it is inevitable that we still detect rRNA sites. Conveniently, this allows us to use the rRNA region as an internal control to ensure the RibOxi-seq chemistry worked.

Figure 18. New Benzonase digestion optimization for RibOxi-seq.

Digestion incubation temperature was elevated to room temperature, and incubation time was 20 minutes. The resulting digestion products are now with sizes averaging 35-40bp long.


Lower
36
1000
800
600
400
200
0
Sample Intensity [FU]
25
200
Length (bp)

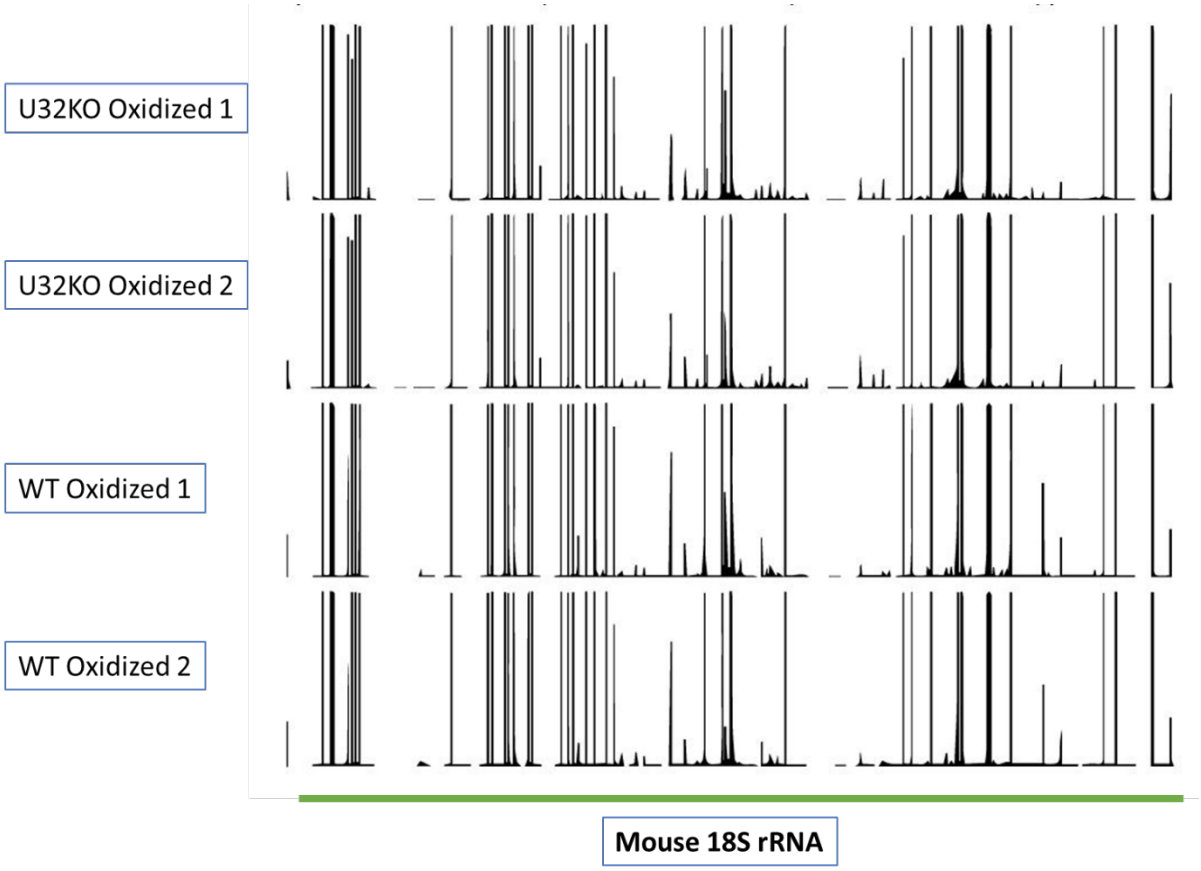
Our results show that experiments on both mouse and human RNAs are largely consistent with known sites in the rRNA region (Figure 19). However, when we examine the data for the rest of the transcriptome, we discovered several issues. First, we witnessed variable mis-priming events, where observed sites have upstream 'motifs' that have 3-8bp matching sequence with the RT primer sequence (Figure 20). Our bioinformatic capabilities did not allow us to filter these sites out quickly as these are of variable length and some contain regions of base pair mismatching. As a consequence, we manually scanned through the genome on the UCSC genome browser and recorded non-mis-priming sites. Another type of seeming artifact is the occurrence of peak doublets with certain lengths of spacing. When examining UCSC tracks for H9 data between a total RNA experiment and a poly(A)$^+$ RNA experiment, we discovered that the poly(A)$^+$ RNA experiment detected almost all known sites in the rRNA region. However, unlike total RNA data, the poly(A)$^+$ RNA data set also displayed additional peak doublets. When we manually filtered out these peaks, total RNA and poly(A)$^+$ RNA data sets became consistent with each other.

Such phenomena suggested that any site that is close to or part of similar doublet peaks are best ignored. In the end, we observed several sites that appear to be real sites. Our final list of sites in H9 cell corresponds to abundantly expressed genes such as desterin, NOMO family genes (NODAL modulators) etc. (Table 9); while mouse datasets generated from liver RNA experiment are still under examination, we found genes such as transferrin, a liver specific gene, is one of the promising sites after filtering (Figure 21).

Figure 19. RibOxi-seq on poly(A)+ enriched mouse liver RNAs and human H9 cell line RNAs.

Top panel is the visual representation of mouse liver 2'-OMe alignment against mm10 genome after performing RibOxi-seq. results between biological replicates are highly consistent with each other. By comparing our site detection with snOPY database reveals that our method can confidently detect mouse 2'-OMe sites (http://snoopy.med.miyazaki-u.ac.jp/).

Bottom panel. RibOxi-seq alignment visualization for H9 RNA 2'-OMe 18S rRNA sites.


U32KO Oxidized 1
U32KO Oxidized 2
WT Oxidized 1
WT Oxidized 2
Mouse 18S rRNA

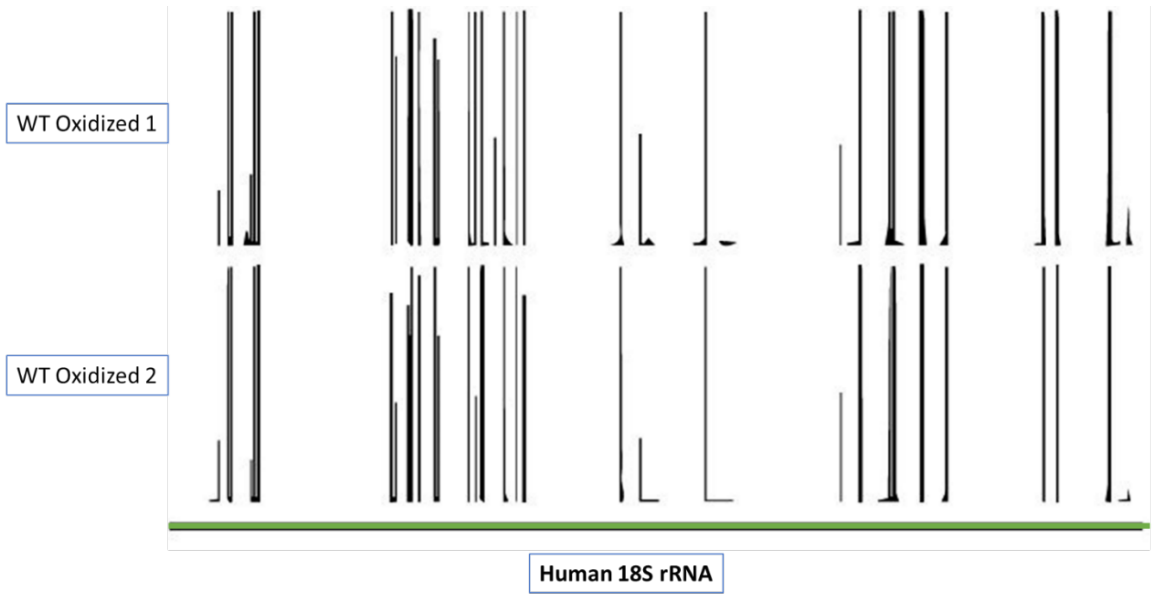

WT Oxidized 1
WT Oxidized 2
Human 18S rRNA

Figure 20. Common artefact occurred during RibOxiseq for transcriptomic 2'-OMe mapping.

Post $NaIO_4$ oxidation only a fraction of terminally methylated RNA fragments remains intact, who are then being ligated to a 3'- linker. The ligated linker act as an anchor point for subsequent reverse-transcription. Due to extremely low abundance of RNA fragment and relatively high concentration of primer oligoes, primer can anneal non-specifically and generate RT products. In the figure, the relative position of such artefact is shown.

RT Primer used in library preparation: 5'-/Biosg/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATTGATGGTGCCTACAG-3'

Reverse complement sequence: CTGTAGGCACCA

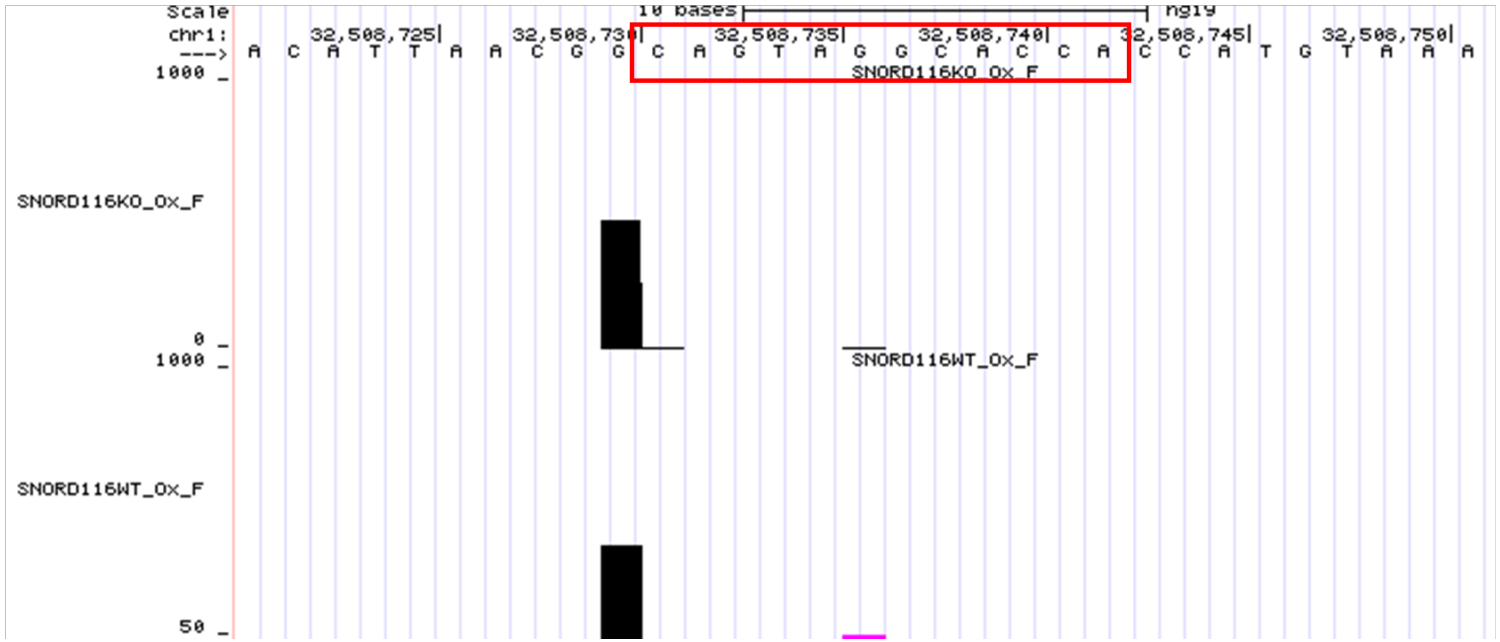
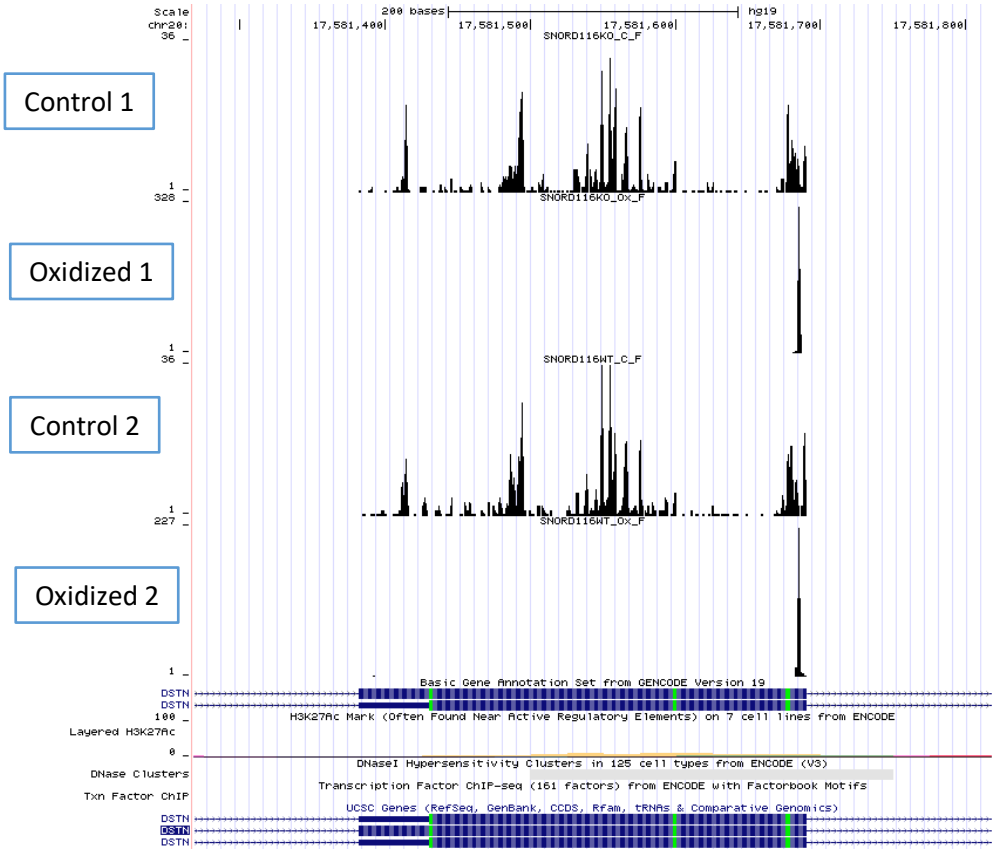
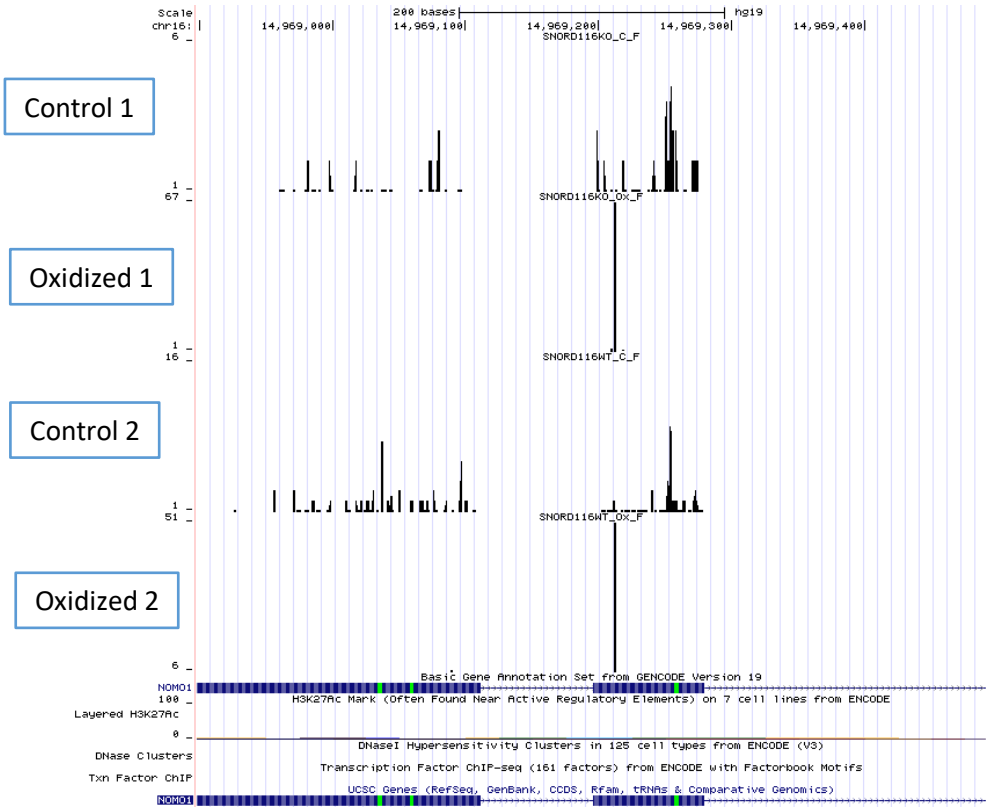


Figure 21. Examples of genes with likely mRNA 2'-OMe sites.

Oxidized samples have promising peaks in NOMO1 and desterin locus. These are not found to be associated with any known chemistry artefact.


Control 1
Oxidized 1
Control 2
Oxidized 2
Scale
chr20:
200 bases
hg19
17,581,400
17,581,500
17,581,600
17,581,700
17,581,800
SNORD116KO_C_F
SNORD116KO_Ox_F
SNORD116WT_C_F
SNORD116WT_Ox_F
Basic Gene Annotation Set from GENCODE Version 19
DSTN
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
Txn Factor ChIP
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)


Control 1
Oxidized 1
Control 2
Oxidized 2
Scale
chr16:
200 bases
hg19
14,969,000
14,969,100
14,969,200
14,969,300
14,969,400
SNORD116KO_C_F
SNORD116KO_Ox_F
SNORD116WT_C_F
SNORD116WT_Ox_F
Basic Gene Annotation Set from GENCODE Version 19
NOMO1
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
Txn Factor ChIP
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)

E. Discussion

In this chapter, we have demonstrated the applicability of the RibOxi-seq method to rRNAs in non-PA1 cell lines and even cells from different organisms. Our results corroborated the evidence that methylation status change of specific *Trypanosome* rRNA 2'-OMe sites between two life stages might be regulated individually. However, how and why such differential patterns occur will require further investigation. One possible way for organisms like mammals is that snoRNAs spliced from introns of host genes can be regulated depending on the host gene expression levels, thus achieving cell type-specific snoRNA expression levels and ultimately cell type -specific 2'-OMe levels. However, there is no evidence of snoRNAs processed from introns of *T. brucei* genes. Then how are snoRNAs of *T. brucei* regulated? Since *T. brucei* snoRNAs are transcribed and processed into polycistronic transcripts before maturation, it is likely pertaining to post-transcriptional regulation (reviewed in Siegel et al. 2011).

In trying to adapt RibOxi-seq to poly(A)$^+$ RNAs, we have learned that the protocol likely needs to be further tweaked and optimized. It was encouraging to show that we were able to detect rRNA sites even with poly(A)$^+$ RNA isolation performed, indicating RibOxi-seq is more flexible than originally anticipated in terms of the amount of required starting materials. Based on such observations, we thus updated the current working protocol. However, mapping sites on mRNAs proved to be more complicated than expected. Although rRNA regions worked well as internal controls, potential artifacts and low read counts in the rest of the transcriptomic regions in non-oxidized samples makes it difficult control for non-specific sites. Our original goal of comparing 2'-OMe sites transcriptome-wide on the WT and SNORD116 snoRNA KO samples we have are therefore intractable at this moment. We will need to further test and make adaptations to the method in order to achieve such capability.

Chapter IV

Conclusions and Future Directions

In summary, we have described in this thesis of successful development of a novel high-throughput method for detection of 2'-OMe in rRNAs - RibOxi-seq. In addition, we have demonstrated wide applicability of the RibOxi-seq through experiments on multiple human cell lines, mouse liver cells, and two *T.brucei* cell types. Through performing RibOxi-seq on human cell lines, we showed that rRNA transcripts within the same cell type are not modified identically (Refer to explanation in chapter 2 and Table 3). We also showed that a decrease in snoRNA levels might not contribute to significant changes in methylation levels. By applying RibOxi-seq to *T.brucei* RNAs and cross confirmation with collaborations, we corroborated the evidence that methylation patterns could be life cycle or developmental stage specific. In trying to adapt the methodology for transcriptome-wide detection, we realized that how the abundance of the transcripts remains a hurdle even when purified poly(A)$^+$ RNAs were used. However, after efforts to filter out possible artifacts, we produced a list of sites that could prove to be real sites, suggesting that we could still achieve transcriptome-wide detection after further optimization of each step of the protocol. Overall, the patterns that our candidate sites correspond to abundantly expressed genes in any particular cell type indicates the method might be restricted to transcripts that are highly expressed. To follow up on the results described in this thesis, two immediate aims will be proposed for future investigation of functions of the 2'-O-methylation.

1. rRNA 2'-OMe landscape profiling of primary cells involved in diseases such as cancers.

Are there associations between rRNA methylation abnormalities and diseases?

rRNAs are most heavily 2'-O-methylated RNA transcripts. Because methylation patterns were

shown to vary between different cell types, we speculate such process forms a layer of regulation, and thus mis-regulation of such process can contribute to pathogenesis (Figure 17B, Table 6, Krogh et al. 2016). To test this possibility, we would choose a cancer type, such as melanoma. The rationale is that cancer cells are highly proliferative, thus requiring high ribosomal capacity, where 2'-OMe mis-regulation is likely to occur. We would like to extract total RNA from primary melanoma cells and matching wild type cells and perform RibOxi-seq. To achieve statistical validity, we reason that we should obtain maximum number of samples as long as cost and sample source allow. The rRNA 2'-OMe sites will be analyzed and compiled for examination of whether there are common depletions or enrichments of modifications at certain sites.

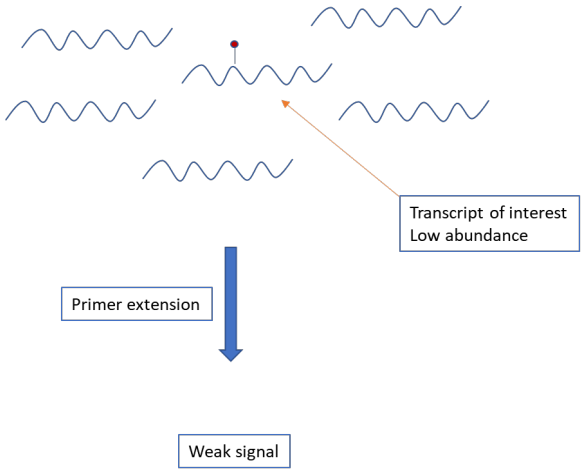
2. Strategies to properly adapt RibOxiseq to transcriptome-wide application.

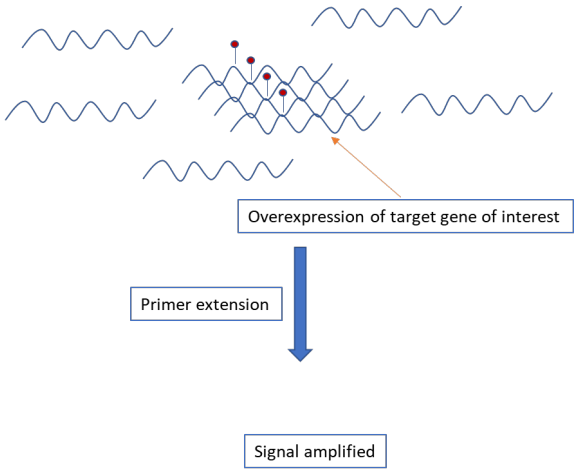
Do currently identified mRNA 2'-OMe sites have merit? Validation of 2'-OMe sites on low abundance RNAs have been historically difficult as discussed in the previous chapter. Primer extension is in general a very sensitive method even for low abundance RNAs, however, we do not know the characteristics of primer extension under low dNTP concentration is (Raymond 2005). Although, PCR amplification coupled with primer extension has been used in several occasions, it is undesirable as it can introduce further noise into the visualization (Holley et al. 2015). Conveniently, the seemingly high efficiency of the snoRNP machinery suggests that it can have much higher methylation capacity than the normal biological requirement for ribosome biogenesis (Figure 22, Newton et al. 2003). Thus, we reason that if we overexpress genes that are on our list such as desterin or NOMO1, and perform primer extension with gene specific primers, we should see corresponding bands if those sites are indeed real. It is also important to determine if these potential methylation sites are guided by any expressed

snoRNAs. As small RNA sequencing technology matured over the past years, it is now possible to accurately profile box C/D snoRNA expression levels through CLIP-seq with antibody to NOP58 protein (Gumienny et al. 2017). Subsequently, the surrounding sequences of the 2'-OMe sites can be bioinformatically analyzed against CLIP-seq snoRNA, from the same cellular source, results to determine the guide RNAs.

Figure 22. Alternative strategy for validating low abundance RNA 2'-OMe.

Overexpression plasmids are transcribed by RNA Pol II, and the overexpressed transcripts undergo same paths as endogenous mRNAs. Thus, if endogenous transcripts are modified with 2'-OMe, the overexpressed copies, at least a significant amount of them, will have 2'-OMe. This can help with detection signal to noise ratio.


Transcript of interest
Low abundance
Primer extension
Weak signal


Overexpression of target gene of interest
Primer extension
Signal amplified

Protocol optimizations for the Illumina platform. So far, reaction composition for each enzymatic step of the RibOxiseq-protocol after oxidation follows manufacturer's instructions with RNA quantity estimates made from the original sample. Thus, we will perform RibOxi-seq library preparation steps until first ligation. We will then use the entire sample to quantify RNA fragment average size and molarity to determine optimal reaction condition. We will extend such strategy to cover all remaining steps. By using such strategy, we should be able to minimize significant amounts of potential artifacts of RibOxi-seq.

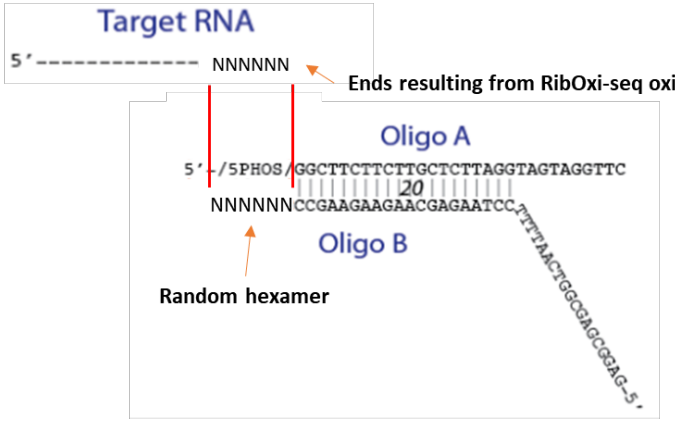
Modify RibOxi-seq to work with Oxford nanopore direct RNA sequencing. Recently, the Minion sequencing platform became widely available. One of the breakthroughs made by this platform is the capability of direct RNA sequencing (Garalde et al. 2018). As a consequence, the library preparation step consists of just two ligations, removing the necessity of conversion to cDNA and PCR amplification. We reason that by adapting RibOxi-seq to the Oxford nanopore platform, we would greatly reduce bias and experimental artifact. In addition, the method can also become quantitative as the only major variable is ligation efficiency. We will design custom oligos so that they can only be ligated to 2'-O methylated ends but not dialdehydes. The oligos can be easily designed to be compatible with the new sequencing platform. In principle, there appear to be no technical barriers for this adaptation, and we will investigate its validity (Figure 23).

Figure 23. Strategy to adapt RibOxi-seq to Oxford Nanopore direct RNA sequencing.

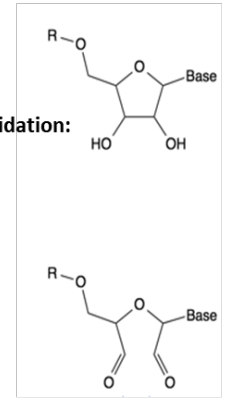
The direct RNA sequencing protocol requires two simple ligation steps for conversion of RNA of interest into RNA library. The second ligation step is universal, while only the first ligation is target specific. The RNA adapter denoted Oligo A/B duplex can be modified to accommodate random annealing and ligation instead of AAAAAAAAAA to UUUUUUUUUU annealing.

Ligation required by official Oxford Nanopore protocol

Adapting Nanopore protocol to RibOxi-seq by designing Adapter oligos



Ends resulting from RibOxi-seq oxidation:



Can be ligated

Cannot be ligated

Table 6. List of reagents and equipment used in RibOxi-seq

Reagents

- Seal-Rite 2.0 mL microcentrifuge tube, natural (USA Scientific, 1620-2720)
- PA-1 [PA1] cell line (ATCC, CRL-1572)
- PureLink RNA Mini Kit (Thermo Fisher Scientific, 12183025)
- PureLink DNase Set (Thermo Fisher Scientific, 12185010)
- TURBO DNA-free Kit (Thermo Fisher Scientific, AM1907)
- Ultra-pure Benzonase (Sigma, E826305KU)
- 10× Benzonase buffer (store at 4°C)
- 3 M sodium acetate pH = 5.2
- Ethanol 100%
- Ethanol 70%
- UltraPure Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v) (Thermo Fisher Scientific, 15593031)
- Acid-Phenol:Chloroform, pH 4.5 (with IAA, 125:24:1) (Thermo Fisher Scientific, AM9720)
- NucAway Spin Columns (Thermo Fisher Scientific, AM10070)
- RNA Analysis ScreenTape (Agilent, 5067-5576)
- RNA Analysis ScreenTape reagents (Agilent, 5067-5577)
- High sensitivity D1000 DNA ScreenTape (Agilent, 5067-5584)
- High sensitivity D1000 DNA ScreenTape reagents (Agilent, 5067-5585)
- Linear polyacrylamide 10 µg/µL (Mullins Molecular Retrovirology Lab Short protocol)
- Sodium meta-periodate (Sigma-Aldrich, 7790-28-5)
- Sodium periodate oxidation buffer: 4.375 mM sodium borate, 50 mM boric acid, pH = 8.6
- 0.2 mL PCR 8-tube FLEX-FREE strip, attached clear flat caps, natural (USA Scientific, 1402-4700)
- β-elimination buffer: 33.75 mM sodium borate, 50 mM boric acid, pH = 9.5
- T4 Polynucleotide Kinase (NEB, M0201L)
- SUPERase• In RNase Inhibitor 20 U/µL (Thermo Fisher Scientific, AM2696)
- RNaseOUT Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific, 10777019)
- T4 RNA Ligase 2, truncated KQ (NEB, M0373S)
- T4 RNA Ligase 1 (NEB, M0204S)
- DMSO 100%
- SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific, 18080051)
- Sodium hydroxide 1 N
- EB buffer: 10 mM Tris-Cl, pH 8.5
- Q5 High-Fidelity 2× Master Mix (NEB, M0492S)
- Agencourt AMPure XP, 450 mL (Beckman Coulter Life Sciences, A63882)
- Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Q32854)
- NextSeq 500/550 Mid Output v2 Kit (150 cycles) (Illumina, FC-404-2001)
- Shrimp alkaline phosphatase (rSAP, NEB, M0371S)

Equipment
Table top centrifuge
Programmable thermal cycler
Heat blocks/water baths
NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, ND-2000)
2200 TapeStation (Agilent Technologies, G2964AA)
Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Q32866)
NextSeq 550 System (Illumina)

Table 7. List of primers and other oligos used.

RibOxi-seq Primers and linkers:

3′ Preadenylated DNA linker (NEB Universal miRNA Cloning Linker, S1315S, dissolve into 50 µM)

- 5′-/rApp/CTGTAGGCACCATCAAT/NH2/-3′

5′ RNA linker (50 µM stock)

- 5′-/Biosg/rArCrArCrGrArCrGrCrUrCrUrUrCrCrGrArUrCrU-3

Reverse transcription primer (50 µM stock)

- 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTRANATTGATGGTGCCTACAG-3′

(Important: The "RAN" represents a customizable random hexamer sequence that can be used to remove PCR duplicates later in the data analysis. Random hexamer sequences are used in the experiments, but longer is recommended for higher sequencing depth.)

Illumina compatible barcoded PCR primers

- i5: 5′-aatgatacggcgaccaccgagatctacacBARCODEacactctttccctacacgacgctcttccgatct-3′
- i7: 5′-caagcagaagacggcatacgagatBARCODEgtgactggagttcagacgtgtgctcttccgatct-3′

(Barcodes are customizable. The protocol is established using paired-end sequencing with dual barcodes. Important: When demultiplexing, the i7 barcode sequence needs to be specified as reverse complement to what is in BARCODE).

It is possible to design longer linker sequences, matching RT primer and PCR primers to increase amplification specificity and efficiency.

Primers for cDNA validation:

Forward:

28S-1 GCTCTCCCACCCCTCCTC

28S-2 CGCAGGTGCAGATCTTGGT

Reverse:

RibOxi-seq RT Primer w/o hexamer:

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT ATTGATGGTGCCTACAG

RT primer sequences for radioactive dNTP concentration dependent primer extension:

LSU C1880 (positive control):	5'-ATGGCCACCGTCCTGCT-3'
SSU U1668 (not detected):	5'-ATCCGAGGGCCTCACTA-3'
LSU A3717 (Novel):	5'-GGCATTTGGCTACCTTA-3'

Anti-sense oligo (ASO) for U63 snoRNA KD:

U63_1:

mA*mG*mU*mU*mU*T*C*C*A*C*A*C*G*T*T*mC*mU*mU*mU*mC

U63_2:

mC*mU*mC*mA*mG*T*C*A*T*T*A*G*T*T*T*mU*mC*mC*mA*mC

Non-specific Control:

Proprietary scrambled LNA oligo from Exiqon

Primers for U63KD qPCR:

Forward:

ATCATTCTGAAAGAACGTGTGG

Reverse:

Universal reverse primer from miScript II RT PCR kit

Table 8. List of 2'-OMe sites detected in *T. brucei* using RibOxiseq

5.8S					
Modification	Position	Predictions	Primer extension	RibOxi-seq PF	RibOxi-seq BF
Um	7	TB10Cs1C3	N/A	No	No
Am	41		N/A	Yes	Yes
Am	43	TB10Cs7C3a/b	N/A	Yes	Yes
Gm	75	TB8Cs3C3	N/A	Yes	Yes
Am	163	TB9Cs4C2	N/A	Yes	Yes

18S					
Modification	Position	Predictions	Primer extension	RibOxi-seq PF	RibOxi-seq BF
Cm	16	TB10Cs2C1	N/A	No	No
Um	36	TB10Cs2C1	N/A	No	No
Cm	46	TB8Cs3C3	N/A	No	No
Am	56	TB8Cs2C1	N/A	No	No
Um	57		N/A	Yes	Yes
Cm	66	TB8Cs2C1	N/A	Yes	Yes
Um	123		N/A	Yes	Yes
Am	125	TB8Cs1C2	N/A	Yes	Yes
Um	386		N/A	Yes	Yes
Cm	507		Yes	No	No
Cm	668		N/A	Yes	Yes
Um	680	TB10Cs3C3	Yes	Yes	Yes
Um	714	TB9Cs2C6	N/A	Yes	Yes
Am	721	TB6Cs2C1	N/A	Yes	Yes
Am	738		N/A	Yes	Yes
Am	818		N/A	No	Yes
Cm	819		N/A	Yes	Yes
Um	977		Yes	No	No
Um	996		Yes	No	No
Um	999		Yes	No	No
Um	1002		Yes	No	No
Um	1038		N/A	Yes	Yes
Um	1425		N/A	Yes	Yes
Gm	1432		N/A	Yes	Yes
Um	1433		N/A	Yes	Yes
Um	1447		N/A	Yes	Yes
Gm	1517	TB11Cs3C2	N/A	Yes	Yes
Gm	1531	TB3Cs1C1	N/A	Yes	Yes
Gm	1532	TB3Cs1C1	N/A	No	No
Gm	1603	TB8Cs3C2	Yes	Yes	Yes
Gm	1628		N/A	Yes	Yes

Um	1630	TB10Cs2"C3	N/A	No	No
Um	1652	TB10Cs1C3	Yes	Yes	Yes
Um	1674	TB10Cs3C2	N/A	Yes	Yes
Gm	1675		N/A	Yes	Yes
Gm	1676	TB9Cs3C1	N/A	Yes	Yes
Gm	1678		N/A	Yes	Yes
Am	1680		N/A	Yes	Yes
Gm	1700	TB10Cs2"C1	N/A	Yes	Yes
Cm	1758		N/A	Yes	Yes
Um	1843		N/A	Yes	Yes
Um	1844	TB7Cs1C1	Yes	Yes	Yes
Am	1871		N/A	Yes	Yes
Cm	1887		Yes	No	No
Gm	1895	TB9Cs2C4	N/A	Yes	Yes
Um	1899	TB11Cs2C1	Yes	Yes	Yes
Gm	1931	TB10Cs4C4	N/A	Yes	Yes
Cm	1932	TB10Cs4C2	N/A	Yes	Yes
Am	2016		Yes	No	No
Am	2050		Yes	No	No
Um	2054	TB8Cs2C0	N/A	Yes	Yes
Cm	2089		Yes	No	No
Am	2096	TB9Cs3C2	Yes	Yes	Yes
Cm	2104		Yes	No	No
Um	2122		N/A	Yes	Yes
Um	2123	TB10Cs4C3	N/A	Yes	Yes
Cm	2134	TB10Cs4C3	Yes	Yes	Yes
Um	2154		N/A	Yes	Yes
Gm	2156		N/A	Yes	Yes
Cm	2216		Yes	No	No
Gm	2227	TB8Cs1C1	N/A	Yes	Yes

28S Alpha

Modification	Position	Predictions	Primer extension	RibOxi-seq PF	RibOxi-seq BF
Am	254	TB10Cs3C2	N/A	Yes	Yes
Um	306		N/A	Yes	Yes
Am	713		N/A	Yes	Yes
Am	743	TB9Cs2C2	N/A	Yes	Yes
Am	744		N/A	Yes	Yes
Am	746	TB10Cs2C2	N/A	Yes	Yes
Cm	760	TB9Cs2C2	N/A	Yes	Yes
Am	762		N/A	Yes	Yes
Gm	880		Yes	Yes	Yes
Cm	887		N/A	No	Yes

Cm	912		N/A	Yes	Yes
Um	914	TB9Cs2C3	N/A	No	No
Gm	915		Yes	No	No
Um	916	TB11Cs4C2	Yes	Yes	No
Gm	925	TB9Cs2C3	N/A	No	No
Am	927	TB11Cs4C2	Yes	Yes	Yes
Cm	970		N/A	Yes	Yes
Gm	976		N/A	No	Yes
Am	986	TB10Cs4C5	N/A	No	No
Gm	992		N/A	Yes	Yes
Am	996	TB9Cs5C2	N/A	Yes	Yes
Gm	998		N/A	Yes	Yes
Um	999		N/A	Yes	Yes
Cm	1006	TB9Cs5C1	N/A	Yes	Yes
Um	1019		N/A	Yes	Yes
Am	1024	TB11Cs1C3	N/A	Yes	Yes
Gm	1027		N/A	Yes	Yes
Gm	1028	TB9Cs2C5	N/A	Yes	Yes
Um	1145	TB10Cs3C4	Yes	Yes	Yes
Um	1166		N/A	Yes	Yes
Am	1179		N/A	Yes	Yes
Am	1180	TB9Cs5C1	N/A	Yes	Yes
Um	1181	TB10Cs3C4	Yes	Yes	Yes
Um	1218		Yes	No	No
Um	1229		N/A	Yes	Yes
Gm	1267	TB9Cs4C1	N/A	Yes	Yes
Um	1330	TB10Cs5C1	N/A	Yes	Yes
Gm	1369		Yes	No	No
Am	1379		Yes	No	No
Am	1391		Yes	No	No
Um	1406		Yes	No	No
Um	1448	TB8Cs2C2	N/A	Yes	Yes
Gm	1572		N/A	No	Yes
Gm	1605	TB10Cs2"C3	N/A	Yes	No
Cm	1608	TB8Cs3C1	Yes	Yes	No
Am	1620	TB8Cs3C1	Yes	Yes	No
Gm	1621		N/A	Yes	Yes
Gm	1634	TB8Cs1C3	Yes	No	No
Gm	1709	TB8Cs1C1	Yes	Yes	Yes
Um	1742	TB7Cs2C1	N/A	Yes	No

28S Beta					
Modification	Position	Prediction	Primer extension	RibOxi-seq PF	RibOxi-seq BF

Gm	71	TB9Cs1C1	N/A	No	No
Um	73		N/A	Yes	Yes
Am	95	TB3Cs3C1	N/A	Yes	Yes
Cm	377	TB9Cs4C3	Yes	Yes	Yes
Um	391		Yes	No	No
Am	395		N/A	Yes	Yes
Am	400	TB6Cs1C2	Yes	Yes	Yes
Um	418	TB11Cs2C1	Yes	No	No
Am	439		N/A	Yes	Yes
Am	520	TB10Cs7C2	N/A	Yes	Yes
Am	544		N/A	Yes	No
Am	545	TB6Cs1C3	Yes	Yes	Yes
Um	546		N/A	Yes	Yes
Um	550		N/A	Yes	Yes
Gm	552	TB10Cs1C4	Yes	Yes	Yes
Um	578	TB11Cs3C1	Yes	Yes	Yes
Am	588	TB8Cs1C3	N/A	Yes	No
Am	590		N/A	Yes	Yes
Um	596		N/A	Yes	Yes
Cm	601	TB10Cs1C1	Yes	Yes	No
Am	609	TB10Cs1C1	Yes	Yes	Yes
Am	621		N/A	Yes	Yes
Am	622	TB10Cs2'C1	Yes	Yes	Yes
Am	646	TB11Cs4'C1	N/A	Yes	Yes
Gm	659	TB9Cs2C7	N/A	Yes	Yes
Um	672		Yes	No	No
Gm	673	TB11Cs4C3	N/A	No	No
Um	685	TB9Cs2C7	Yes	Yes	Yes
Um	719	TB10Cs3C5	N/A	No	No
Um	728	TB10Cs3C5	N/A	Yes	No
Um	729		N/A	Yes	Yes
Um	931		N/A	Yes	No
Cm	969		N/A	No	Yes
Um	1011	TB6Cs1C2	N/A	No	No
Gm	1019		N/A	No	Yes
Gm	1062	TB11Cs2C1	Yes	Yes	Yes
Gm	1094	TB5Cs1C1	N/A	Yes	Yes
Um	1098		N/A	Yes	Yes
Cm	1175	TB5Cs1C1	N/A	Yes	Yes
Am	1201	TB10Cs2"C2	N/A	Yes	Yes
Gm	1245	TB3Cs1C-1	Yes	Yes	Yes
Gm	1247	TB11Cs1C2	Yes	Yes	Yes
Cm	1264	TB6Cs1C1	N/A	Yes	No

Gm	1269	TB11Cs1C2	Yes	Yes	Yes
Gm	1293		Yes	No	No
Gm	1295		Yes	No	No
Cm	1333		Yes	No	No
Um	1336		Yes	No	No
Gm	1349		Yes	No	Yes
Um	1375	TB10Cs3C1	Yes	Yes	No
Gm	1376		N/A	Yes	Yes
Am	1388	TB11Cs4C1	Yes	Yes	Yes
Am	1400	TB11Cs4C1	Yes	Yes	Yes
Am	1412		N/A	Yes	No
Cm	1413	TB9Cs2C1	N/A	Yes	Yes
Um	1429		N/A	Yes	Yes
Um	1434		N/A	Yes	Yes
Um	1435	TB9Cs3C3	N/A	Yes	Yes

Sites detection is summarized from 5.8S, 18S, 28S alpha and 28S beta rRNAs. Detected sites are compared with bioinformatic predictions and previous experimental data from Michaeli lab and Michaeli et al. (2012).

Table 9. List of potential H9 mRNA sites:

APLP2	Amyloid Beta Precursor Like Protein 2
GLTSCR2	NOP53 Ribosome Biogenesis Factor
HIGD1A	Hypoxia-Inducible Gene 1 Protein
HNRNPA2B1	Heterogeneous Nuclear Ribonucleoprotein A2/B1
NBR1	Neighbor Of BRCA1 Gene 1
NOMO2	NODAL Modulator 2
NOMO3	NODAL Modulator 3
PRMT1	Protein Arginine Methyltransferase 1
RPL21	Ribosomal Protein L21
U2AF2	U2 Small Nuclear RNA Auxiliary Factor 2
ZMYM3	Zinc Finger MYM-Type Containing 3

V. Bibliography

Alefelder, S., Patel, B. K., & Eckstein, F. (1998). Incorporation of terminal phosphorothioates into oligonucleotides. Nucleic acids research, 26(21), 4983-4988.

Amort, T., Rieder, D., Wille, A., Khokhlova-Cubberley, D., Riml, C., Trixl, L., … Lusser, A. (2017). Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. Genome Biology, 18(1), 1–16.

Barth, S., Shalem, B., Hury, A., Tkacz, I. D., Liang, X. H., Uliel, S., … Michaeli, S. (2008). Elucidating the role of C/D snoRNA in rRNA processing and modification in Trypanosoma brucei. Eukaryotic Cell, 7(1), 86–101.

Baserga, S. J., Yang, X. D., & Steitz, J. A. (1991). An intact Box C sequence in the U3 snRNA is required for binding of fibrillarin, the protein common to the major family of nucleolar snRNPs. The EMBO Journal, 10(9), 2645–51.

Basse, C., & Arock, M. (2015). The increasing roles of epigenetics in breast cancer: Implications for pathogenicity, biomarkers, prevention and treatment. International Journal of Cancer.

Begley, C. G., Lipkowitz, S., Göbel, V., Mahon, K. a, Bertness, V., Green, a R., … Kirsch, I. R. (1992). Molecular characterization of NSCL, a gene encoding a helix-loop-helix protein expressed in the developing nervous system. Proceedings of the National Academy of Sciences of the United States of America, 89(1), 38–42.

Bhalla, T., Rosenthal, J. J. C., Holmgren, M., & Reenan, R. (2004). Control of human potassium channel inactivation by editing of a small mRNA hairpin. Nature Structural and Molecular Biology, 11(10), 950–956.

Bieth, E., Eddiry, S., Gaston, V., Lorenzini, F., Buffet, A., Conte Auriol, F., … Tauber, M. (2015). Highly restricted deletion of the SNORD116 region is implicated in Prader-Willi Syndrome. European Journal of Human Genetics: EJHG, 23(2), 252–5.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & development, 16(1), 6-21.

Boccaletto, P., MacHnicka, M. A., Purta, E., Pitkowski, P., Baginski, B., Wirecki, T. K., … Bujnicki, J. M. (2018). MODOMICS: A database of RNA modification pathways. 2017 update. Nucleic Acids Research, 46(D1), D303–D307.

Bokar, J. A., Shambaugh, M. E., Polayes, D., Matera, A. G., & Rottman, F. M. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA (New York, N.Y.), 3(11), 1233–47.

Bortolin-Cavaillé, M.-L., & Cavaillé, J. (2012). The SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) gene clusters at the imprinted Prader-Willi locus generate canonical box C/D snoRNAs. Nucleic Acids Research, 40(14), 6800–7.

Bourgeois, G., Marcoux, J., Saliou, J. M., Cianférani, S., & Graille, M. (2017). Activation mode of the eukaryotic m2G10 tRNA methyltransferase Trm11 by its partner protein Trm112. Nucleic Acids Research, 45(4), 1971–1982.

Brameier, M., Herwig, A., Reinhardt, R., Walter, L., & Gruber, J. (2011). Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. Nucleic Acids Research, 39(2), 675–686.

Brown, L., Espinosa, R., Le Beau, M. M., Siciliano, M. J., & Baer, R. (1992). HEN1 and HEN2: a subgroup of basic helix-loop-helix genes that are coexpressed in a human neuroblastoma. Proceedings of the National Academy of Sciences of the United States of America, 89(18), 8492–6.

Burgess, A. L., David, R., & Searle, I. R. (2015). Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae. BMC Plant Biology, 15(1), 1–17.

Cameron, V., & Uhlenbeck, O. C. (1977). 3′-Phosphatase Activity in T4 Polynucleotide Kinase. Biochemistry, 16(23), 5120–5126.

Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., ... & Agris, P. F. (2011). The RNA modification database, RNAMDB: 2011 update. Nucleic acids research, 39(suppl 1), D195-D201.

Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., & Gilbert, W. V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature, 515(7525), 143–146.

Cassidy, S. B., Schwartz, S., Miller, J. L., & Driscoll, D. J. (2012). Prader-Willi syndrome. Genetics in Medicine, 14(1), 10–26.

Cavaillé J, Bachellerie JP. 1998. SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction. Nucleic Acids Res 26: 1576–1587.

Cavaillé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C. I., Horsthemke, B., … Hüttenhofer, A. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proceedings of the National Academy of Sciences of the United States of America, 97(26), 14311–6.

Chamberlain, S. J., Chen, P., Ng, K. Y., Bourgois-rocha, F., Lemtiri-chlieh, F., Levine, E. S., & Lalande, M. (2010). Induced pluripotent stem cell models of the genomic imprinting disorders Angelman and Prader – Willi syndromes. Pnas, 107(41), 17668–17673.

Chen, L.-L., & Carmichael, G. G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Molecular Cell, 35(4), 467–78.

Cheng, J. X., Chen, L., Li, Y., Cloe, A., Yue, M., Wei, J., … Vardiman, J. W. (2018). RNA cytosine methylation and methyltransferases mediate chromatin organization and 5-azacytidine response and resistance in leukaemia. Nature Communications, 9(1), 1163.

Chikne, V., Doniger, T., Rajan, K. S., Bartok, O., Eliaz, D., Cohen-Chalamish, S., … Michaeli, S. (2016). A pseudouridylation switch in rRNA is implicated in ribosome function during the life cycle of Trypanosoma brucei. Scientific Reports, 6(April), 25296.

Choi, J., Ieong, K.-W., Demirci, H., Chen, J., Petrov, A., Prabhakar, A., … Puglisi, J. D. (2016). N6-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. Nature Structural & Molecular Biology, 23(2), 110–115.

Clouet-D'Orval, B., Gaspin, C., & Mougin, A. (2005). Two different mechanisms for tRNA ribose methylation in Archaea: A short survey. In Biochimie (Vol. 87, pp. 889–895).

Cohn, W. E., & Volkin, E. (1951). Nucleoside-5′-phosphates from ribonucleic acid. Nature, 167(4247), 483-484.

Correns, C. (1937). Non-Mendelian inheritance. Non-Mendelian inheritance.

Cui, Q., Shi, H., Ye, P., Li, L., Qu, Q., Sun, G., … Shi, Y. (2017). m6A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells. Cell Reports, 18(11), 2622–2634.

Cuzin, F. (2013). Non-Mendelian Inheritance, In Brenner's Encyclopedia of Genetics (Second Edition). Academic Press, 96-97

Daffis, S., Szretter, K. J., Schriewer, J., Li, J., Youn, S., Errett, J., … Diamond, M. S. (2010). 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. Nature, 468(7322), 452–456.

Dai, Q., Moshitch-Moshkovitz, S., Han, D., Kol, N., Amariglio, N., Rechavi, G., … He, C. (2017). Nm-seq maps 2′-O-methylation sites in human mRNA with base precision. Nature Methods, 14(7), 695–698.

Decatur, W. A., & Fournier, M. J. (2002). rRNA modifications and ribosome function. Trends in Biochemical Sciences.

Dennis, P. P., Tripp, V., Lui, L., Lowe, T., & Randau, L. (2015). C/D box sRNA-guided 2'-O-methylation patterns of archaeal rRNA molecules. BMC Genomics, 16(1), 632.

Ding, F., Li, H. H., Zhang, S., Solomon, N. M., Camper, S. a, Cohen, P., & Francke, U. (2008). SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. PloS One, 3(3), e1709.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., … Rechavi, G. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature, 485(7397), 201–206.

Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M. S., … He, C. (2016). The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. Nature, 530(7591), 441–446.

Dong, H., Chang, D. C., Hua, M. H. C., Lim, S. P., Chionh, Y. H., Hia, F., … Shi, P. Y. (2012). 2'-O methylation of internal adenosine by flavivirus NS5 methyltransferase. PLoS Pathogens, 8(4).

Dunn, D.B. (1961) Biochim. Biophys. Acta, 46, 198-200.

Ender, C., Krek, A., Friedländer, M. R., Beitzinger, M., Weinmann, L., Chen, W., … Meister, G. (2008). A Human snoRNA with MicroRNA-Like Functions. Molecular Cell, 32(4), 519–528.

Esguerra, J., Warringer, J., & Blomberg, A. (2008). Functional importance of individual rRNA 2'-O-ribose methylation revealed by high-resolution phenotyping. Rna, 14(Tollervey 1987), 649–656.

Felsenfeld, G. (2007). A Brief History of Epigenetics - Chapter 2.pdf. Epigenetics, 15–22.

Filipowicz W, Pogaci c V (2002) Biogenesis of small nucleolar ribonucleoproteins. Curr Opin Cell Biol 14(3):319–327.

Fukuda, M., Umeno, H., Nose, K., Nishitarumizu, A., Noguchi, R., & Nakagawa, H. (2017). Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-To-I RNA editing. Scientific Reports, 7(December 2016), 8–19.

Furuichi,Y., Morgan,M., Shatkin,A.J., Jelinek,W., Salditt- Georgieff,M. and Darnell,J.E. (1975) Methylated, blocked 5 termini in HeLa cell mRNA. Proc. Natl Acad. Sci. USA, 72, 1904–1908.

Galardi, S., Fatica, A., & Bachi, A. (2002). Purified Box C/D snoRNPs Are Able To Reproduce Site-Specific 2'-O-Methylation of Target RNA In Vitro. … and Cellular Biology, 22(19), 6663–6668.

Ganot, P., Jady, B. E., Bortolin, M. L., Darzacq, X., & Kiss, T. (1999). Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. Molecular and Cellular Biology, 19(10), 6906–6917.

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., … Turner, D. J. (2018). Highly parallel direct RN A sequencing on an array of nanopores. Nature Methods, 15(3), 201–206.

Gardner, K. E., Allis, C. D., & Strahl, B. D. (2011). Operating on chromatin, a colorful language where context matters. Journal of molecular biology, 409(1), 36-46.

Garren, S. B., Kondaveeti, Y., Duff, M. O., & Carmichael, G. G. (2015). Global Analysis of Mouse Polyomavirus Infection Reveals Dynamic Regulation of Viral and Host Gene Expression and Promiscuous Viral RNA Editing. PLoS Pathogens, 11(9), 1–25.

Ge, J., & Yu, Y. T. (2013). RNA pseudouridylation: new insights into an old modification. Trends in biochemical sciences, 38(4), 210-218.

Ge, J., Liu, H., & Yu, Y. T. (2010). Regulation of pre-mRNA splicing in Xenopus oocytes by targeted 2'-O-methylation. RNA, 16(5), 1078–1085.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-Divon, M., … Hanna, J. H. (2015). m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. Science, 347(6225), 1002–1006.

Gillen, A. E., Yamamoto, T. M., Kline, E., Hesselberth, J. R., & Kabos, P. (2016). Improvements to the HITS-CLIP protocol eliminate widespread mispriming artifacts. BMC Genomics, 17(1), 1–11.

Greger, I. H., Khatri, L., & Ziff, E. B. (2002). RNA editing at Arg607 controls AMPA receptor exit from the endoplasmic reticulum. Neuron, 34(5), 759–772.

Greger, I. H., Khatri, L., Kong, X., & Ziff, E. B. (2003). AMPA receptor tetramerization is mediated by Q/R editing. Neuron, 40(4), 763–774.

Gu, R., Zhang, Z., DeCerbo, J. N., & Carmichael, G. G. (2009). Gene regulation by sense-antisense overlap of polyadenylation signals. RNA, 15(6), 1154–1163.

Gumienny, R., Jedlinski, D. J., Schmidt, A., Gypas, F., Martin, G., Vina-vilaseca, A., & Zavolan, M. (2017). High-throughput identification of C / D box snoRNA targets with CLIP and RiboMeth-seq, 45(5), 2341–2353.

Guy, M. P., Podyma, B. M., Preston, M. A., Shaheen, H. H., Krivos, K. L., Limbach, P. A., … Phizicky, E. M. (2012). Yeast Trm7 interacts with distinct proteins for critical modifications of the tRNAPhe anticodon loop. RNA, 18(10), 1921–1933.

Guy, M. P., Shaw, M., Weiner, C. L., Hobson, L., Stark, Z., Rose, K., … Phizicky, E. M. (2015). Defects in tRNA Anticodon Loop 2′-O-Methylation Are Implicated in Nonsyndromic X-Linked Intellectual Disability due to Mutations in FTSJ1. Human Mutation, 36(12), 1176–1187.

He, C. (2010). Grand Challenge Commentary: RNA epigenetics? Nature Chemical Biology, 6(12), 863–865.

Higuchi, M., Single, F. N., Köhler, M., Sommer, B., Sprengel, R., & Seeburg, P. H. (1993). RNA editing of AMPA receptor subunit GluR-B: A base-paired intron-exon structure determines position and efficiency. Cell, 75(7), 1361–1370.

Hirose, T., Shu, M. Di, & Steitz, J. A. (2003). Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. Molecular Cell, 12(1), 113–123.

Hocine, S., Singer, R. H., & Grünwald, D. (2010). RNA processing and export. Cold Spring Harbor Perspectives in Biology, 2(12), a000752.

Hocine, S., Singer, R. H., & Grünwald, D. (2010). RNA processing and export. Cold Spring Harbor Perspectives in Biology, 2(12), a000752.

Holley, C. L., Li, M. W., Scruggs, B. S., Matkovich, S. J., Ory, D. S., & Schaffer, J. E. (2015). Cytosolic accumulation of small nucleolar RNAs (snoRNAs) is dynamically regulated by NADPH oxidase. Journal of Biological Chemistry, 290(18), 11741–11748.

Honda S, Morichika K, Kirino Y. 2016. Selective amplification and sequencing of cyclic phosphate-containing RNAs by the cP-RNA-seq method. Nat Protoc 11: 476–489.

Hyde, J. L., & Diamond, M. S. (2015). Innate immune restriction and antagonism of viral RNA lacking 2′-O methylation. Virology, 479–480(2), 66–74.

Incarnato, D., Anselmi, F., Morandi, E., Neri, F., Maldotti, M., Rapelli, S., … Oliviero, S. (2016). High-throughput single-base resolution mapping of RNA 2'-O-methylated residues. Nucleic Acids Research, 45(3), gkw810.

Jády, B. E., & Kiss, T. (2000). Characterisation of the U83 and U84 small nucleolar RNAs: two novel 2'-O-ribose methylation guide-RNAs that lack complementarities to ribosomal RNAs. Nucleic Acids Research. 28 (6): 1348–1354.

Jády, B. E., & Kiss, T. (2001). A small nucleolar guide RNA functions both in 2′-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. EMBO Journal, 20(3), 541–551.

Johannsen, W. (1905). Arvelighedslærens elementer: forelæsninger holdte ved Københavns universitet. Gyldendal.

Johnson, K. C., Houseman, E. A., King, J. E., Von Herrmann, K. M., Fadul, C. E., & Christensen, B. C. (2016). 5-Hydroxymethylcytosine localizes to enhancer elements and is associated with survival in glioblastoma patients. Nature Communications, 7, 1–11.

Jorjani, H., Kehr, S., Jedlinski, D. J., Gumienny, R., Hertel, J., Stadler, P. F., … Gruber, A. R. (2016). An updated human snoRNAome. Nucleic Acids Research, gkw386.

Jung, S., von Thülen, T., Laukemper, V., Pigisch, S., Hangel, D., Wagner, H., … Bauer, S. (2015). A single naturally occurring 2'-O-methylation converts a TLR7- and TLR8-activating RNA into a TLR8-specific ligand. PloS One, 10(3), e0120498.

Karijolich, J., & Yu, Y.-T. (2007). Spliceosomal snRNA modifications and their function. RNA Biology, 7(2), 192–204.

Kirino, Y., & Mourelatos, Z. (2007). Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. Nature Structural and Molecular Biology, 14(4), 347–348.

Kirino, Y., & Mourelatos, Z. (2007). The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. RNA (New York, N.Y.), 13(9), 1397–401.

Kishore, S., & Stamm, S. (2006). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. Science (New York, N.Y.), 311(5758), 230–2.

Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. The EMBO Journal, 20(14), 3617–22.

Kiss, T., Fayet, E., Jády, B. E., Richard, P., & Weber, M. (2006). Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. Cold Spring Harbor Symposia on Quantitative Biology, 71, 407–17.

Kiss-László, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M., & Kiss, T. (1996). Site-specific ribose methylation of preribosomal RNA: A novel function for small nucleolar RNAs. Cell, 85(7), 1077–1088.

Kowles R. (2001) Non-Mendelian Inheritance. In: Solving Problems in Genetics. Springer

Krogh N, Jansson MD, Häfner SJ, Tehler D, Birkedal U, Christensen-Dalsgaard M, Lund AH, Nielsen H. 2016. Profiling of 2′-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. Nucleic Acids Res 44: 7884–7895.

Kurata, S., Ohtsuki, T., Suzuki, T., & Watanabe, K. (2003). Quick two-step RNA ligation employing periodate oxidation. Nucleic Acids Research, 31(22), e145–e145.

Lafontaine, D. L. J. (2015). Noncoding RNAs in eukaryotic ribosome biogenesis and function. Nature Structural & Molecular Biology, 22(1), 11–19.

Lee, J., Harris, A. N., Holley, C. L., Mahadevan, J., Pyles, K. D., Lavagnino, Z., … Schaffer, J. E. (2016). Rpl13a small nucleolar RNAs regulate systemic glucose metabolism. Journal of Clinical Investigation, 126(12), 4616–4625.

Lence, T., Akhtar, J., Bayer, M., Schmid, K., Spindler, L., Ho, C. H., … Roignant, J. Y. (2016). M6A modulates neuronal functions and sex determination in Drosophila. Nature, 540(7632), 242–247.

Lestrade, L. (2006). snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nucleic Acids Research, 34(90001), D158–D162.

Levesque-Sergerie, J. P., Duquette, M., Thibault, C., Delbecchi, L., & Bissonnette, N. (2007). Detection limits of several commercial reverse transcriptase enzymes: Impact on the low- and high-abundance transcript levels assessed by quantitative RT-PCR. BMC Molecular Biology, 8.

Li, R., & Fox, A. H. (2016). SPArking Interest in the Long Noncoding RNA World: A New Class of 5′ SnoRNA-Stabilized LncRNA that Influences Alternative Splicing. Molecular Cell, 64(3), 435–437.

Li, X., Ma, S., & Yi, C. (2016). Pseudouridine: the fifth RNA nucleotide with renewed interests. Current Opinion in Chemical Biology, 33(Figure 2), 108–116.

Liang, X.-H., Uliel, S., Hury, A., Barth, S., Doniger, T., Unger, R., & Michaeli, S. (2005). A genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in Trypanosoma brucei reveals a trypanosome-specific pattern of rRNA modification. RNA (New York, N.Y.), 11(5), 619–45.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15: 550.

Maden, B. E. (2001). Mapping 2'-O-methyl groups in ribosomal RNA. Methods (San Diego, Calif.), 25(3), 374–82.

Maden, B. E. H., Corbett, M. E., Heeney, P. A., Pugh, K., & Ajuh, P. M. (1995). Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. Biochimie, 77(1–2), 22–29.

Marchand, V., Blanloeil-oillo, F., Helm, M., & Motorin, Y. (2016). Illumina-based RiboMethSeq approach for mapping of 2 -O-Me residues in RNA. Nucleic Acids Research, 1–12.

Marz, M., Gruber, A. R., Höner Zu Siederdissen, C., Amman, F., Badelt, S., Bartschat, S., … Stadler, P. F. (2011). Animal snoRNAs and scaRNAs with exceptional structures. RNA Biology, 8(6), 938–946.

Massenet, S., Motorin, Y., Lafontaine, D. L., Hurt, E. C., Grosjean, H., & Branlant, C. (1999). Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. Molecular and cellular biology, 19(3), 2142-2154.

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., & Jaffrey, S. R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons. Cell, 149(7), 1635–1646.

Michaeli, S., Doniger, T., Gupta, S. K., Wurtzel, O., Romano, M., Visnovezky, D., … Ullu, E. (2012). RNA-seq analysis of small RNPs in Trypanosoma brucei reveals a rich repertoire of non-coding RNAs. Nucleic Acids Research.

Michel, C. I., Holley, C. L., Scruggs, B. S., Sidhu, R., Brookheart, R. T., Listenberger, L. L., … Schaffer, J. E. (2011). Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. Cell Metabolism, 14(1), 33–44.

Modifications in and around the anticodon loop of 2011), and mutants lacking C38 and C39 in their tRNAs have tRNA play several crucial roles in translation (Agris et al. increased +1 and ?1 frameshifting (Lecointe et al. 2002; 2007; Phizicky and Hopper 2010).

Motorin, Y. and Grosjean, H. (1998) In Grosjean, H. and Benne, R. (eds.) Modification and Editing of RNA. ASM Press, Washington D.C., 543-9.

Mummaneni, P., & Shord, S. S. (2014). Epigenetics and oncology. Pharmacotherapy.

Munafó, D. B., & Robb, G. B. (2010). Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. RNA (New York, N.Y.), 16(12), 2537–2552.

Newton, K., Petfalski, E., Tollervey, D., & Cáceres, J. F. (2003). Fibrillarin is essential for early development and required for accumulation of an intron-encoded small nucleolar RNA in the mouse. Molecular and Cellular Biology, 23(23), 8519–27.

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. Annual Review of Biochemistry, 79(3), 321–49.

Nishikura, K. (2016). A-to-I editing of coding and non-coding RNAs by ADARs. Nature Reviews Molecular Cell Biology, 17(2), 83–96.

Niu, Y., Zhao, X., Wu, Y. S., Li, M. M., Wang, X. J., & Yang, Y. G. (2013). N6-methyl-adenosine (m6A) in RNA: An Old Modification with A Novel Epigenetic Function. Genomics, Proteomics and Bioinformatics, 11(1), 8–17.

O'Connell, M. (2015). RNA modification and the epitranscriptome; the next frontier. Rna, 21(4), 703–704.

O'Connell, M. A., Mannion, N. M., & Keegan, L. P. (2015). The Epitranscriptome and Innate Immunity. PLoS Genetics, 11(12), 1–11.

Oerum, S., Dégut, C., Barraud, P., & Tisné, C. (2017). m1A post-transcriptional modification in tRNAs. Biomolecules, 7(1), 1–15.

Olejniczak, M., Galka, P., & Krzyzosiak, W. J. (2009). Sequence-non-specific effects of RNA interference triggers and microRNA regulators. Nucleic Acids Research, 38(1), 1–16.

Ordog, T., Syed, S. A., Hayashi, Y., & Asuzu, D. T. (2012). Epigenetics and chromatin dynamics: A review and a paradigm for functional disorders. Neurogastroenterology and Motility.

Patil, D. P., Chen, C. K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., & Jaffrey, S. R. (2016). M6A RNA methylation promotes XIST-mediated transcriptional repression. Nature, 537(7620), 369–373.

Patterson, D. G., Roberts, J. T., King, V. M., Houserova, D., Barnhill, E. C., Crucello, A., … Borchert, G. M. (2017). Human snoRNA-93 is processed into a microRNA-like RNA that promotes breast cancer cell invasion. Npj Breast Cancer, 3(1), 25.

Pelletier, J., Thomas, G., & Volarevi, S. (2017). Ribosome biogenesis in cancer: New players and therapeutic avenues. Nature Reviews Cancer, 18(1), 51–63.

Pintard, L., Lecointe, F., Bujnicki, J. M., Bonnerot, C., Grosjean, H., & Lapeyre, B. (2002). Trm7p catalyses the formation of two 2'-O-methylriboses in yeast tRNA anticodon loop. EMBO Journal, 21(7), 1811–1820.

Poison, A. G., McCloskey, J. A., Bass, B. L., Crain, P. F., Pomerantz, S. C., & McCloskey, J. A. (1991). The Mechanism of Adenosine to Inosine Conversion by the Double-Stranded rna Unwinding/Modifying Activity: A High-Performance Liquid Chromatography-Mass Spectrometry Analysis. Biochemistry, 30(49), 11507–11514.

Ponte-Sucre, A. (2016). An overview of trypanosoma brucei infections: An intense host-parasite interaction. Frontiers in Microbiology.

Qiu, F., & McCloskey, J. a. (1999). Selective detection of ribose-methylated nucleotides in RNA by a mass spectrometry-based method. Nucleic Acids Research, 27(18), e20.

Raabe, C. a, Tang, T.-H., Brosius, J., & Rozhdestvensky, T. S. (2014). Biases in small RNA deep sequencing data. Nucleic Acids Research, 42(3), 1414–26.

Raymond, C. K. (2005). Simple, quantitative primer-extension PCR assay for direct monitoring of microRNAs and short-interfering RNAs. RNA, 11(11), 1737–1744.

Renalier, M.-H., Joseph, N., Gaspin, C., Thebault, P., & Mougin, A. (2005). The Cm56 tRNA modification in archaea is catalyzed either by a specific 2'-O-methylase, or a C/D sRNP. RNA (New York, N.Y.), 11(7), 1051–63.

Richard, P., & Kiss, T. (2006). Integrating snoRNP assembly with mRNA biogenesis. EMBO Reports, 7(6), 590–2.

Rimbach, K., Kaiser, S., Helm, M., Dalpke, A. H., & Eigenbrod, T. (2015). 2'-O-Methylation within Bacterial RNA Acts as Suppressor of TLR7/TLR8 Activation in Human Innate Immune Cells. Journal of Innate Immunity, 7, 482–493.

Ritchey, L. E., Su, Z., Tang, Y., Tack, D. C., Assmann, S. M., & Bevilacqua, P. C. (2017). Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure in vivo. Nucleic Acids Research, 45(14), e135.

Ronchetti, D., Mosca, L., Cutrona, G., Tuana, G., Gentile, M., Fabris, S., … Neri, A. (2013). Small nucleolar RNAs as new biomarkers in chronic lymphocytic leukemia. BMC Medical Genomics, 6(1), 27.

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., & Siomi, M. C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes & Development, 21(13), 1603–8.

Saletore, Y., Meyer, K., Korlach, J., Vilfan, I. D., Jaffrey, S., & Mason, C. E. (2012). The birth of the Epitranscriptome: deciphering the function of RNA modifications. Genome biology, 13(10), 175.

Saraiya, A. A., & Wang, C. C. (2008). snoRNA, a Novel Precursor of microRNA in Giardia lamblia. PLoS Pathogens, 4(11), e1000224.

Satoh, A., Takai, K., Ouchi, R., Yokoyama, S., & Takaku, H. (2000). Effects of anticodon 2'-O-methylations on tRNA codon recognition in an Escherichia coli cell-free translation. Rna, 6(5), 680–686.

Schaefer, M., Pollex, T., Hanna, K., & Lyko, F. (2009). RNA cytosine methylation analysis by bisulfite sequencing. Nucleic Acids Research.

Schmitt, F. C. F., Freund, I., Weigand, M. A., Helm, M., Dalpke, A. H., & Eigenbrod, T. (2017). Identification of an optimized 2'-O-methylated trinucleotide RNA motif inhibiting Toll-like receptors 7 and 8. RNA (New York, N.Y.), 23(9), 1344–1351.

Schultz, A., Nottrott, S., Watkins, N. J., & Lührmann, R. (2006). Protein-protein and protein-RNA contacts both contribute to the 15.5 K-mediated assembly of the U4/U6 snRNP and the box C/D snoRNPs. Molecular and cellular biology, 26(13), 5146-5154.

Schwartz, S., Bernstein, D. A., Mumbach, M. R., Jovanovic, M., Herbst, R. H., León-Ricardo, B. X., ... & Fink, G. (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell, 159(1), 148-162.

Scott, M. S., Avolio, F., Ono, M., Lamond, A. I., & Barton, G. J. (2009). Human miRNA precursors with box H/ACA snoRNA features. PLoS Computational Biology, 5(9).

Siegel, T. N., Gunasekera, K., Cross, G. A., & Ochsenreiter, T. (2011). Gene expression in Trypanosoma brucei: lessons from high throughput RNA sequencing studies. Trends Parasitol, 27(10), 434–441.

Simon, B., Kirkpatrick, J. P., Eckhardt, S., Reuter, M., Rocha, E. A., Andrade-Navarro, M. A., … Carlomagno, T. (2011). Recognition of 2′-o-methylated 3′-end of piRNA by the PAZ domain of a Piwi protein. Structure, 19(2), 172–180.

Siprashvili, Z., Webster, D. E., Johnston, D., Shenoy, R. M., Ungewickell, A. J., Bhaduri, A., … Khavari, P. A. (2015). The noncoding RNAs SNORD50A and SNORD50B bind K-Ras and are recurrently deleted in human cancer. Nature Genetics, 48(1), 53–58.

Sloan, K. E., Warda, A. S., Sharma, S., Entian, K. D., Lafontaine, D. L. J., & Bohnsack, M. T. (2017). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. RNA Biology, 14(9), 1138–1152.

Smietanski, M., Werner, M., Purta, E., Kaminska, K. H., Stepinski, J., Darzynkiewicz, E., … Bujnicki, J. M. (2014). Structural analysis of human 2'-O-ribose methyltransferases involved in mRNA cap structure formation. Nature Communications, 5, 3004.

Somme, J., Van Laer, B., Roovers, M., Steyaert, J., Versées, W., & Droogmans, L. (2014). Characterization of two homologous 2′-O-methyltransferases showing different specificities for their tRNA substrates. Rna, 20(8), 1257–1271.

Starr, J.L. and Fefferman, R. (1964) J. Biol. Chem., 239, 3457-3461.

Stepanov, G. A., Filippova, J. A., Komissarov, A. B., Kuligina, E. V, Richter, V. A., & Semenov, D. V. (2015). Regulatory role of small nucleolar RNAs in human diseases. Biomed Res Int, 2015, 206849.

Tanaka-Fujita, R., Soeno, Y., Satoh, H., Nakamura, Y., & Mori, S. (2007). Human and mouse protein-noncoding snoRNA host genes with dissimilar nucleotide sequences show chromosomal synteny. RNA (New York, N.Y.), 13(6), 811–6.

Tkaczuk, K. L., Dunin-Horkawicz, S., Purta, E., & Bujnicki, J. M. (2007). Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. BMC Bioinformatics, 8.

Tollervey, D., Lehtonen, H., Jansen, R., Kern, H., & Hurt, E. C. (1993). Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. Cell, 72(3), 443–457.

Tycowski, K. T., Smith, C. M., Shu, M. D., & Steitz, J. A. (1996). A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in Xenopus. Proceedings of the National Academy of Sciences of the United States of America, 93(25), 14480–5.

Uliel, S., Liang, X. H., Unger, R., & Michaeli, S. (2004). Small nucleolar RNAs that guide modification in trypanosomatids: Repertoire, targets, genome organisation, and unique functions. International Journal for Parasitology, 34(4), 445–454.

van Heyningen, V., & Yeyati, P. L. (2004). Mechanisms of non-Mendelian inheritance in genetic disease. Human molecular genetics, 13(suppl_2), R225-R233.

van Nues RW, Granneman S, Kudla G, Sloan KE, Chicken M, Tollervey D, Watkins NJ. 2011. Box C/D snoRNP catalysed methylation is aided by additional pre-rRNA base-pairing. EMBO J 30: 2420–2430.

Vitali, P., Basyuk, E., Le Meur, E., Bertrand, E., Muscatelli, F., Cavaillé, J., & Huttenhofer, A. (2005). ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. Journal of Cell Biology, 169(5), 745–753.

Waddington, C. H. (1942). The epigenotype. International Journal of Epidemiology, 41(1), 10–13.

Walsh, D., & Mohr, I. (2011). Viral subversion of the host protein synthesis machinery. Nature Reviews Microbiology, 9(12), 860–875.

Wang, X., & He, C. (2014). Dynamic RNA modifications in posttranscriptional regulation. Molecular Cell, 56(1), 5–12. http://doi.org/10.1016/j.molcel.2014.09.001

Watkins NJ, Bohnsack MT (2012) The box C/D and H/ACA snoRNPs: Key players in the modification, processing and the dynamic folding of ribosomal RNA. Wiley Interdiscip Rev RNA 3(3):397–414. 4.

Wei, C., Xiao, R., Chen, L., Cui, H., Zhou, Y., Xue, Y., … Fu, X. D. (2016). RBFox2 Binds Nascent RNA to Globally Regulate Polycomb Complex 2 Targeting in Mammalian Genomes. Molecular Cell, 62(6), 875–889.

Werner, M., Purta, E., Kaminska, K. H., Cymerman, I. A., Campbell, D. A., Mittra, B., … Bujnicki, J. M. (2011). 2′-O-ribose methylation of cap2 in human: Function and evolution in a horizontally mobile family. Nucleic Acids Research, 39(11), 4756–4768.

Wu, H., Yin, Q. F., Luo, Z., Yao, R. W., Zheng, C. C., Zhang, J., … Chen, L. L. (2016). Unusual Processing Generates SPA LncRNAs that Sequester Multiple RNA Binding Proteins. Molecular Cell, 64(3), 534–548.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., … Yang, Y.-G. (2016). Nuclear m 6 A Reader YTHDC1 Regulates mRNA Splicing. Molecular Cell, 61(4), 507–519.

Xue, S., & Barna, M. (2012). Specialized ribosomes: A new frontier in gene regulation and organismal biology. Nature Reviews Molecular Cell Biology, 13(6), 355–369.

Yang, Z., Lin, J., & Ye, K. (2016). Box C/D guide RNAs recognize a maximum of 10 nt of substrates. PNAS, 201604872.

Yin, Q.-F., Yang, L., Zhang, Y., Xiang, J.-F., Wu, Y.-W., Carmichael, G. G., & Chen, L.-L. (2012). Long noncoding RNAs with snoRNA ends. Molecular Cell, 48(2), 219–30.

Yoshihama, M., Nakao, A., & Kenmochi, N. (2013). snOPY: a small nucleolar RNA orthological gene database. BMC Research Notes, 6, 426.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant mi- croRNA biogenesis. Science 307: 932–935.

Yu, Y. T., Shu, M. D., & Steitz, J. A. (1997). A new method for detecting sites of 2'-O-methylation in RNA molecules. RNA (New York, N.Y.), 3(3), 324–31.

Yuki, H., Yamaji, A., & Fujiwara, S. (1976). [Analysis of 5-methylcytosine in tRNA by gas chromatography (author's transl)]. Yakugaku Zasshi : Journal of the Pharmaceutical Society of Japan, 96(10), 1184–8.

Zhao, B. S., Roundtree, I. A., & He, C. (2016). Post-transcriptional gene regulation by mRNA modifications. Nature Reviews Molecular Cell Biology, 18(1), 31–42.

Zhuang, F., Fuchs, R. T., Sun, Z., Zheng, Y., & Robb, G. B. (2012). Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. Nucleic Acids Research, 40(7), 1–14.

Züst, R., Cervantes-Barragan, L., Habjan, M., Maier, R., Neuman, B. W., Ziebuhr, J., … Thiel, V. (2011). Ribose 2′-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. Nature Immunology, 12(2), 137–143.